

UNITED STATES PATENT & TRADEMARK OFFICE

Re: Application of: **Frank MELTKE**

Serial No.: To Be Assigned

Filed: Herewith as national phase of International
Application No. PCT/EP2004/002915, filed 19 March 2004

For: **SEARCH SYSTEM AND METHOD FOR
DETERMINING INFORMATION FROM A
DATABANK, ESPECIALLY THE WORDWIDE WEB**

Customer No: 23280

LETTER RE: PRIORITY

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

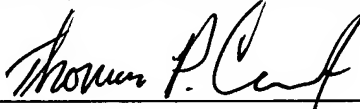
September 22, 2005

Sir:

Applicant hereby claims priority of German Application Serial No. 103 13 420.4, filed 25
March 2003, through International Application No. PCT/EP2004/002915, filed 19 March 2004.

Respectfully submitted,

DAVIDSON, DAVIDSON & KAPPEL, LLC

By 
Thomas P. Canty
Reg. No. 44,586

Davidson, Davidson & Kappel, LLC
485 Seventh Avenue, 14th Floor
New York, New York 10018
(212) 736-1940

BEST AVAILABLE COPY



**Prioritätsbescheinigung über die Einreichung
einer Patentanmeldung**

Aktenzeichen: 103 13 420.4

Anmeldetag: 25. März 2003

Anmelder/Inhaber: Frank M e l t k e , 24852 Eggebek/DE

Bezeichnung: Recherchensystem und Verfahren zur Ermittlung
von Informationen aus einer Datenbank, insbesonde-
re aus dem World-Wide-Web

IPC: G 06 F 17/30

Die angehefteten Stücke sind eine richtige und genaue Wiedergabe der ur-
sprünglichen Unterlagen dieser Patentanmeldung.

München, den 29. März 2004
Deutsches Patent- und Markenamt
Der Präsident
Im Auftrag

Stanschus

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

P030300DE-6/20

25. März 2003

Ansprüche

1. Recherchensystem (1) zur Ermittlung von Informationen aus einer Datenbank unter Nutzung einer Mehrzahl von Suchmaschinen (2) mit einem Integrationsmodul (6), das die von den Suchmaschinen (2) gelieferten, jeweils eine Mehrzahl von Ergebnisdatensätzen (E) umfassenden Ergebnisdaten in ein gemeinsames Datenformat konvertiert und anschließend unter Rückgriff auf in einem Speichermodul (8) hinterlegte Priorisierungsparameter (P) jedem Ergebnisdatensatz (E) einen Relevanzkennwert zuordnet.
2. Recherchensystem (1) nach Anspruch 1, bei dem die Priorisierungsparameter (P) abhängig von der jeweiligen Suchmaschine (2) gewählt sind.
3. Recherchensystem (1) nach Anspruch 1 oder 2, dessen Integrationsmodul (6) den Relevanzkennwert für den jeweiligen Ergebnisdatensatz (E) unter Berücksichtigung eines von der jeweiligen Suchmaschine (2) bereitgestellten Ausgangsrelevanzkennwerts ermittelt.
4. Recherchensystem (1) nach einem der Ansprüche 1 bis 3, dessen Integrationsmodul (6) über ein gemeinsames Pufferspeichermodul (12) mit einem Distributionsmodul (10) kommuniziert, das einen im gemeinsamen Datenformat eingehenden Suchauftrag für jede ausgewählte Suchmaschine (2) in deren spezifisches Datenformat umwandelt und anschließend an diese ausgibt.
5. Recherchensystem (1) nach Anspruch 4, dessen Distributionsmodul (10) mit einem zweiten Speichermodul (14) verbunden ist, in dem Syntaxdaten für die Suchmaschinen (2) hinterlegt sind.

6. Recherchensystem (1) nach Anspruch 4 oder 5, in dessen Pufferspeichermodul (12) für einzelne Suchaufträge charakteristische Identifikationsdaten hinterlegbar sind.
7. Recherchensystem (1) nach einem der Ansprüche 4 bis 6, dessen Distributionsmodul (10) eingangsseitig über ein Filtermodul (16) mit einem Ein-/Ausgabemodul (18) verbunden ist.
8. Recherchensystem (1) nach einem der Ansprüche 4 bis 7, dessen Distributionsmodul (10) mit dem Integrationsmodul (6) in eine gemeinsame Einheit integriert ist.
9. Recherchensystem (1) nach einem der Ansprüche 4 bis 8, bei dem das Distributions- und das Integrationsmodul (6) datenseitig mit einer Werbemitteldatenbank (20) verbunden sind, die als Antwortdatensatz auf eine eingehende Suchanfrage einen hinsichtlich eines Kennungssignals an die Suchanfrage angepassten Datensatz liefert.
10. Verfahren zur Ermittlung von Informationen aus einer Datenbank, insbesondere aus dem World-Wide-Web, unter Nutzung eines Recherchensystems (1) nach einem der Ansprüche 1 bis 9, bei dem die von den Suchmaschinen (2) gelieferten, jeweils eine Mehrzahl von Ergebnisdatensätzen (E) umfassenden Ergebnisdaten in ein gemeinsames Datenformat konvertiert werden, wobei anschließend unter Rückgriff auf in einem Speichermodul (8) hinterlegte Priorisierungsparameter (P) jedem Ergebnisdatensatz (E) einen Relevanzkennwert zugeordnet wird.
11. Verfahren nach Anspruch 10, bei dem die Priorisierungsparameter (P) abhängig von der jeweiligen Suchmaschine (2) gewählt werden.
12. Verfahren nach Anspruch 10 oder 11, bei dem der Relevanzkennwert für den jeweiligen Ergebnisdatensatz (E) unter Berücksichtigung eines von der

jeweiligen Suchmaschine (2) bereitgestellten Ausgangsrelevanzkennwerts ermittelt wird.

13. Verfahren nach einem der Ansprüche 10 bis 12, bei dem ein im gemeinsamen Datenformat eingehender Suchauftrag für jede ausgewählte Suchmaschine (2) in deren spezifisches Datenformat umgewandelt und anschließend an diese ausgegeben wird.

14. Verfahren nach einem der Ansprüche 10 bis 13, bei dem ein über ein Ein-/Ausgabemodul (18) vorgegebener Suchauftrag vor seiner weiteren Bearbeitung einer Filterung unterzogen wird.

15. Verfahren nach einem der Ansprüche 10 bis 14, bei dem ein Suchauftrag zusätzlich an eine Werbemitteldatenbank (20) ausgegeben wird, wobei ein von dieser eingehender Antwortdatensatz mittels eines Kennungssignals der jeweiligen Suchanfrage zugeordnet wird.

Beschreibung

Recherchensystem und Verfahren zur Ermittlung von Informationen aus einer Datenbank, insbesondere aus dem World-Wide-Web

Die Erfindung bezieht sich auf ein Recherchensystem zur Ermittlung von Informationen aus einer Datenbank unter Nutzung einer Mehrzahl von Suchmaschinen. Sie betrifft weiter ein Verfahren zur Ermittlung von Informationen aus einer Datenbank, insbesondere aus dem World-Wide-Web, unter Nutzung eines derartigen Recherchensystems.

In komplexen Datenbanken oder auch im weltumspannenden Computernetz („World-Wide-Web“) wird eine enorme Menge an Informationen bereitgehalten, die ein Benutzer zu Recherchezwecken mehr oder weniger gezielt abrufen kann. Um dabei Informationen aus der großen Informationsmenge heraus gezielt nutzbar machen zu können, kommen so genannte Suchmaschinen zum Einsatz, die gerade bei der Informationsbeschaffung aus dem World-Wide-Web teilweise erhebliche Verbreitung erfahren haben. Bei derartigen Suchmaschinen wird dem Benutzer üblicherweise über ein Ein-/Ausgabemodul ein Anfragenfenster zur Verfügung gestellt, über das gezielte Such- oder Recherchebegriffe vorgegeben werden können. Anschließend durchsucht die Suchmaschine die Informationsbasis der Datenbank bzw. des World-Wide-Web nach geeigneten Schlüssel- oder Keywords. Die daraufhin aufgefundenen Antwortdatensätze werden üblicherweise von der jeweiligen Suchmaschine hinsichtlich ihrer Relevanz für den vorgegebenen Suchauftrag kategorisiert und dem Benutzer in der Art einer Trefferliste in nach ihrer Relevanz geordneter Reihenfolge zur Verfügung gestellt.

Zwischenzeitlich haben sich eine Vielzahl derartiger Suchmaschinen etabliert, die sich hinsichtlich ihrer Leistungsfähigkeit, Zuverlässigkeit und/oder der berücksichtigten Informationsbasis in der Art einer thematischen Spezialisierung voneinander unterscheiden können. Gerade weil aber aufgrund der Vielzahl der beispielsweise im World-Wide-Web angebotenen Informationen beim Betrieb einer derartigen

1 Suchmaschine eine gewisse Schwerpunktbildung oder Spezialisierung einzelner
2 Suchmaschinen erforderlich sein kann, kann die Qualität der Suchergebnisse ab-
3 hängig von der vorgegebenen Suchanfrage für einzelne Suchmaschinen in gewis-
4 sen Bandbreiten variieren.

5
6 Der Erfindung liegt daher die Aufgabe zugrunde, ein Recherchensystem zur Er-
7 mittlung von Informationen aus einer Datenbank, insbesondere aus dem World-
8 Wide-Web, anzugeben, das zur Erzielung einer gleichbleibend hohen Recherche-
9 qualität im besonderem Maße zur gemeinsamen und/oder gleichzeitigen Nutzung
10 einer Mehrzahl von Suchmaschinen besonders geeignet ist. Des Weiteren soll ein
11 Verfahren zur Ermittlung von Informationen aus einer Datenbank, insbesondere
12 aus dem World-Wide-Web, unter Nutzung eines derartigen Recherchensystems
13 angegeben werden.

14
15 Bezüglich des Recherchensystems wird diese Aufgabe erfindungsgemäß gelöst
16 mit einem Integrationsmodul, das die von den Suchmaschinen gelieferten, jeweils
17 eine Mehrzahl von Ergebnisdatensätzen umfassenden Ergebnisdaten in ein ge-
18 meinsames Datenformat konvertiert und anschließend unter Rückgriff auf in ei-
19 nem im Speichermodul hinterlegte Priorisierungsparameter jedem Ergebnisdaten-
20 satz einen Relevanzkennwert zuordnet.

21
22 Die Erfindung geht dabei von der Überlegung aus, dass zur Erreichung gleichblei-
23 bend hoher Recherchequalitäten grundsätzlich eine Mehrzahl von alternativen
24 Suchmaschinen eingesetzt werden sollte, so dass suchfragenabhängige Quali-
25 tätsmängel wechselweise ausgeglichen werden können. Gerade beim Einsatz
26 einer Mehrzahl von Suchmaschinen für eine Suchanfrage ist jedoch damit zu
27 rechnen, dass eine große Vielzahl an Recherchetreffern oder Ergebnisdatensät-
28 zen von der Gesamtheit der Suchmaschinen generiert wird. Für eine benutzer-
29 freundliche und gezielte Auswertung dieser Ergebnisse sollte eine allen Ergebnis-
30 sen gemeinsame Relevanzkennung vergeben werden, so dass in für den Benut-
31 zer nachvollziehbarerweise eine systematische Ordnung sämtlicher Rechercheer-
32 gebnisse nach vorgegebenen Relevanzkriterien erfolgen kann. Insbesondere

1 sollte dabei ermöglicht sein, die nach vorgegebenen Kriterien als relevantest er-
2 achteten Ergebnisse zuerst zu präsentieren, so dass der Benutzer sich auf die
3 wesentlichen Rechercheergebnisse konzentrieren kann. Um dies zu ermöglichen,
4 ist zunächst die Konvertierung der eingehenden Datensätze in ein gemeinsames
5 Datenformat vorgesehen, so dass eine gemeinsame Datenauswertung und -struk-
6 turierung ermöglicht wird. Anschließend ist die Gliederung der aufgefundenen Er-
7 gebnisdatensätze im Sinne einer gemeinsamen Relevanzstruktur vorgesehen.
8 Diese Funktion der Zusammenfügung und Konsolidierung von Ergebnisdatensät-
9 zen aus verschiedenen Quellen wird durch das Integrationsmodul erfüllt.

10
11 Die hierarchische Kategorisierung der aufgefundenen Ergebnisdatensätze durch
12 Zuordnung eines spezifischen Relevanzkennwerts kann dabei nach vom Benutzer
13 oder einem Betreiber vorgegebenen Priorisierungskriterien erfolgen. Insbesondere
14 könnte vorgesehen sein, bei der Auswertung der gelieferten Ergebnisse bestimm-
15 ten Suchmaschinen generell und standardisiert eine erhöhte Relevanz zuzuord-
16 nen, beispielsweise weil für einen bestimmten Typus einer Suchanfrage bekannt
17 ist, dass bestimmte Suchmaschinen in der Regel vergleichsweise hohe Recher-
18 chequalität liefern. Um dies zu ermöglichen, sind die Priorisierungsparameter vor-
19 teilhafterweise abhängig von der jeweiligen Suchmaschine gewählt, so dass bei-
20 spielsweise anhand des für eine bestimmte Suchmaschine hinterlegten Priorisie-
21 rungskennwerts festgestellt werden kann, dass ein von dieser Suchmaschine ge-
22 lieferte Ergebnisdatensatz mit einem vergleichsweise hohen Relevanzkennwert
23 versehen wird.

24
25 Die Vorgabe einer spezifischen Priorisierung bei der Berücksichtigung der Such-
26 maschinen kann dabei vom Benutzer nach Art und Umfang vorgegeben sein. In-
27 nerhalb der von einer Suchmaschine gelieferten Ergebnisdatensätze kann die so
28 genannte interne Priorisierung anschließend auf der Grundlage der von der Such-
29 maschine bereits gelieferten Priorisierung erfolgen. Dazu wird in weiterer oder al-
30 ternativer vorteilhafter Ausgestaltung der Relevanzkennwert für den jeweiligen
31 Ergebnisdatensatz unter Berücksichtigung eines von der jeweiligen Suchmaschine
32 bereitgestellten Ausgangsrelevanzkennwerts vom Integrationsmodul ermittelt.

1
2 Dabei kann insbesondere in der Art einer „dynamischen Priorisierung“ eine Nor-
3 malisierung von den jeweiligen Suchmaschinen bereits mitgelieferter Ausgangs-
4 relevanzkennwerte vorgesehen sein, bei der der von der jeweiligen Suchmaschine
5 als relevantest angesehene Ergebnisdatensatz mit einem vorläufigen Relevanz-
6 kennwert von beispielsweise 100 % versehen wird, wobei die von der jeweiligen
7 Suchmaschine bereits als vergleichsweise weniger relevant eingestuft weiteren
8 Ergebnisdatensätze mit einem entsprechend abgestuften Relevanzkennwert ver-
9 sehen werden. Bei dieser Art der Auswertung und Priorisierung ist somit in der Art
10 einer Gleichberechtigung sämtlicher eingesetzter Suchmaschinen ein direkter
11 Quervergleich einzelner Ergebnisdatensätze ermöglicht. Zusätzlich oder alternativ
12 kann aber auch vorgesehen sein, dass bereits bei der Normalisierung, also bei
13 der Bildung der Bezugsgröße für den von der jeweiligen Suchmaschine als rele-
14 vantest angesehenen Ergebnisdatensatz, ein suchmaschinenspezifischer Kenn-
15 wert mit vorgegeben wird, durch den der Relevanzkennwert des von der jeweiligen
16 Suchmaschine als relevantest angesehenen Ergebnisdatensatzes suchmaschi-
17 nenspezifisch vorgegeben wird. Durch eine derartige suchmaschinenspezifische
18 Vorgabe sind somit Relativwichtungen zwischen einzelnen Suchmaschinen mög-
19 lich, in die insbesondere auch Erfahrungswerte über die generelle Zuverlässigkeit
20 der jeweiligen Suchmaschine und/oder auch themen- oder suchanfragenabhängi-
21 ge Erkenntnisse über die jeweilige Suchmaschine mit einfließen können.

22
23 Das Integrationsmodul ist zur gemeinsamen Auswertung der von den verschiede-
24 nen Suchmaschinen gelieferten Ergebnisdatensätzen ausgelegt, wobei zunächst
25 eine Konvertierung der einzelnen Ergebnisdatensätze in ein gemeinsames und
26 somit gemeinsam auswertbares Datenformat vorgesehen ist. Dazu ist das Inte-
27 grationsmodul dafür ausgelegt, die spezifische, von der jeweiligen Suchmaschine
28 eingesetzte Syntax oder deren Datenformat auswerten und gegebenenfalls um-
29 wandeln zu können. Um dies zu ermöglichen, ist das Integrationsmodul vorteil-
30 hafterweise mit einem zweiten Speichermodul verbunden, in dem Syntaxdaten für
31 die einzelnen Suchmaschinen hinterlegt sind. Anhand dieser Syntaxdaten kann

1 sodann die suchmaschinenspezifische Datenkonvertierung vorgenommen wer-
2 den.

3
4 In weiterer vorteilhafter Ausgestaltung ist zusätzlich zum Integrationsmodul auch
5 ein Distributionsmodul vorgesehen, das einen im gemeinsamen Datenformat ein-
6 gehenden Suchauftrag geeignet auf die einzelnen zum Einsatz vorgesehenen
7 Suchmaschinen aufteilt. Dazu ist das Distributionsmodul, das zweckmäßigerweise
8 über ein gemeinsames Pufferspeichermodule mit dem Integrationsmodul kommu-
9 niziert, vorteilhafterweise dafür ausgelegt, den jeweiligen Suchauftrag geeignet in
10 eine für die jeweilige Suchmaschine verarbeitbare Syntax oder ein entsprechen-
11 des Datenformat umzuwandeln.

12 Um eine zuverlässige Zuordnung der gelieferten Ergebnisse zum jeweiligen Such-
13 auftrag auch dann zu ermöglichen, wenn in enger zeitlicher Reihenfolge eine ver-
14 gleichsweise große Vielzahl an Suchaufträgen zu bearbeiten ist, was insbesonde-
15 re bei unterschiedlichen Reaktionszeiten der Suchmaschine zu unterschiedlichen
16 Wartezeiten auf die Ergebnisse und somit zu zeitlichem Überlapp zwischen ver-
17 schiedenen Suchaufträgen führen könnte, ist dem Integrationsmodul und dem
18 Distributionsmodul vorteilhafterweise ein gemeinsames Pufferspeichermodule zu-
19 geordnet. Im Pufferspeichermodule werden zweckmäßigerweise für jeden Such-
20 auftrag charakteristische Identifikationsdaten, beispielsweise eine Identifikations-
21 kennummer, hinterlegt, wobei die einzelnen, an die Suchmaschinen direkt ausge-
22 gebenen spezifizierten Suchaufträge mit einer entsprechenden Kennnummer ver-
23 sehen sind. Bei den von den Suchmaschinen gelieferten Antworten ist dann im
24 Integrationsmodul durch einen Rückgriff auf die im Pufferspeicher hinterlegten
25 charakteristischen Identifikationsdaten eine eindeutige Zuordnung der Ergebnis-
26 datensätze zum jeweiligen Suchauftrag ermöglicht, so dass eine konsolidierte und
27 konsistente Auswertung der eingehenden Ergebnisse ermöglicht ist.
28
29

30 Das Distributionsmodul ist vorteilhafterweise spezifisch dafür ausgelegt, den ein-
31 gehenden Suchauftrag datenseitig in eine für die jeweiligen Suchmaschinen ver-
32 arbeitbares Datenformat oder eine Syntax zu konvertieren. Dazu ist auch das Di-

1 distributionsmodul vorteilhafterweise mit dem zweiten Speichermodul verbunden, in
2 dem Syntaxdaten für die Suchmaschinen hinterlegt sind. Durch gezielte Wartung
3 und Pflege des zweiten Speichermoduls ist somit auf besondere Weise ermög-
4 licht, gegebenenfalls Aktualisierungen der von den Suchmaschinen verwendeten
5 Syntax oder auch die Hinzufügung oder Löschung vollständiger Suchmaschinen
6 auf besonders einfache Weise vorzunehmen.

7
8 In weiterer vorteilhafter Ausgestaltung ist das Distributionsmodul eingangsseitig
9 über ein Filtermodul mit einem Ein-/Ausgabemodul verbunden, über das vom Be-
10 nutzer einzelne Suchaufträge vorgebbar sind. Das Filtermodul kann dabei insbe-
11 sondere dafür ausgelegt sein, anhand einstellbarer oder individuell vorgebbarer
12 Filterkriterien bestimmte Suchanfragen ganz zu unterdrücken oder beispielsweise
13 länder- oder kundenspezifische Individualvorgaben hinzuzufügen. Dabei kann
14 insbesondere auch vorgesehen sein, anhand der Art der Anfrage gezielt einzelne,
15 als besonders geeignet erkannte Suchmaschinen für die nachfolgende Bearbei-
16 tung des Suchauftrags vorzugeben. Das Filtermodul kann somit insbesondere
17 auch dafür ausgelegt sein, abhängig von der jeweiligen Suchanfrage oder von de-
18 ren thematischem Schwerpunkt eine Vor- oder Grobauswahl der einzusetzenden
19 Suchmaschinen zu treffen und somit eine Grobstrategie für die nachfolgende
20 Strategie vorzugeben.

21
22 Grundsätzlich können das Distributionsmodul und das Integrationsmodul als ei-
23 genständige Einheiten, beispielsweise auf Software- oder auf Hardwarebasis,
24 ausgeführt sein. Eine besonders kompakte und somit günstige Bauweise ist dabei
25 erreichbar, indem das Distributionsmodul vorteilhafterweise mit dem Integrations-
26 modul in eine gemeinsame Einheit integriert ist.

27
28 Es ist weit verbreitete Praxis, gerade bei der Durchführung von Recherchen im
29 Internet oder World-Wide-Web, dass dem Benutzer gemeinsam mit den auf seine
30 Suchanfrage hin generierten Trefferlisten Werbeanzeigen, so genannte „Ban-
31 ners“, mit angezeigt werden. Dabei werden in Reaktion auf eine Suchanfrage von
32 einem auch als „Ad-Server“ bezeichneten Werbemitteldatenbank gemeinsam mit

1 den Ergebnissen zufällig aus den dort hinterlegten Banners ausgewählte Daten-
2 sätze mit übermittelt, die anschließend auf dem Ein-/Ausgabemodul zur Anzeige
3 gebracht werden. Bei einem derartigen System erfolgt die Hinzufügung der indivi-
4 duellen Banners jedoch üblicherweise unspezifiziert und insbesondere ohne Be-
5 zug auf das vom Benutzer durch die Suchanfrage vorgegebene Thema, so dass
6 das Benutzerinteresse am jeweiligen Werbemittel nur vergleichsweise gering ist.
7 In besonders vorteilhafter Weiterbildung des oben genannten Systems sind je-
8 doch nunmehr das Distributions- und das Integrationsmodul datenseitig mit einer
9 Werbemitteldatenbank verbunden, die als Antwortdatensatz auf eine eingehende
10 Suchanfrage einen hinsichtlich eines Kennungssignals an die Suchanfrage ange-
11 passten Datensatz liefert. Mit anderen Worten: Aufgrund der ohnehin vorgesehe-
12 nen, vergleichsweise qualifizierten Auswertung der vorgegebenen Suchanfrage,
13 die über die entsprechende Konvertierung im Distributionsmodul ermöglicht wird,
14 ist nunmehr die Erzeugung einer beispielsweise inhalts- oder themenbezogenen
15 Kennung für die Suchanfrage ermöglicht, die bei einer Weiterleitung der Suchan-
16 frage auf die Werbemitteldatenbank eine gezielte Selektion von hinsichtlich des
17 Kennungssignals, also beispielsweise hinsichtlich des bearbeiteten Themas, be-
18 sonders geeigneten Werbedatensatzes ermöglicht. Durch eine derartige gezielte
19 Zuordnung der als Antwortsignal mitgegebenen Werbemitteldatensätze ist somit
20 eine deutlich erhöhte Aufmerksamkeit beim Benutzer gegenüber der eingesetzten
21 Werbemittel erreichbar.

22
23 Bezüglich des Verfahrens wird die genannte Aufgabe unter Nutzung des genann-
24 ten Recherchensystems gelöst, indem die von den Suchmaschinen gelieferten,
25 jeweils eine Mehrzahl von Ergebnisdatensätzen umfassenden Ergebnisdaten in
26 ein gemeinsames Datenformat konvertiert werden, wobei anschließend unter
27 Rückgriff auf in einem Speichermodul hinterlegte Priorisierungsparameter jedem
28 Ergebnisdatensatz ein Relevanzkennwert zugeordnet wird. Über das Integrations-
29 modul ist somit vorgesehen, für die gelieferten Ergebnisdatensätze ein konsoli-
30 diertes und konsistentes Relevanzsystem zu erzeugen, in dem die einzelnen Er-
31 gebnisdatensätze relativ zueinander im Hinblick auf vorgegebene Relevanzkriteri-
32 en geeignet positioniert sind.

1
2 - Vorteilhafterweise werden die Priorisierungsparameter dabei abhängig von der
3 jeweiligen Suchmaschine gewählt, wobei der Relevanzkennwert in alternativer
4 oder zusätzlicher vorteilhafter Weiterbildung für den jeweiligen Ergebnisdatensatz
5 unter Berücksichtigung eines von der jeweiligen Suchmaschine bereitgestellten
6 Ausgangsrelevanzkennwerts ermittelt wird.

7
8 Vorteilhafterweise wird ein im gemeinsamen Datenformat eingehender Suchauf-
9 trag für jede ausgewählte Suchmaschine in deren spezifisches Datenformat um-
10 gewandelt und anschließend an diese ausgegeben. In alternativer oder zusätzli-
11 cher vorteilhafter Ausgestaltung wird ein über ein Ein-/Ausgabemodul vorgegebe-
ner Suchauftrag vor seiner weiteren Bearbeitung einer Filterung unterzogen.

13
14 Vorteilhafterweise wird ein Suchauftrag zusätzlich zu den Suchmaschinen an eine
15 Werbemitteldatenbank ausgegeben, wobei ein von dieser eingehender Antwort-
16 datensatz mittels eines Kennungssignals der jeweiligen Suchanfrage zugeordnet
17 wird. Durch das Kennungssignal ist dabei insbesondere ermöglicht, den Antwort-
18 datensatz hinsichtlich vorgebbarer Kriterien, beispielsweise hinsichtlich verwen-
19 deter Inhalte oder Themen, an die jeweilige Suchanfrage anzupassen, so dass im
20 Ergebnis dem Benutzer gemeinsam mit seinen Suchergebnissen ein thematisch
21 zur Suchanfrage passender Werbeträger angeboten wird.

22
24 Die mit der Erfindung erzielten Vorteile bestehen insbesondere darin, dass durch
25 die Verwendung des Integrationsmoduls, dass eingehende Ergebnisdatensätze
26 auf ein gemeinsames Datenformat konvertiert und anschließend eine relative Ge-
27 wichtung und Priorisierung der Ergebnisdatensätze untereinander durch Vergabe
28 eines „globalen“ Relevanzkennwerts eine konsistente gemeinsame Nutzung einer
29 Mehrzahl von Suchmaschinen ermöglicht ist. Durch das mit dem Integrationsmo-
30 dul kommunizierende Distributionsmodul ist weiterhin gewährleistet, dass unter
31 Rückgriff auf im Speichermodul hinterlegte, suchmaschinenspezifische Daten wie
32 beispielsweise Syntaxdaten eine gezielte Ansprache verschiedenartiger Suchma-
schinen mit einer gemeinsamen Suchanfrage ermöglicht ist, wobei ebenso eine

1 konsolidierte gemeinsame Auswertung der von den angesprochenen Suchma-
 2 schinen gelieferten Ergebnisse unter Berücksichtigung von deren spezifischer
 3 Syntax ermöglicht ist.

4
 5 Ein Ausführungsbeispiel der Erfindung wird anhand einer Zeichnung näher erläu-
 6 tert. Darin zeigen:

7
 8 Fig. 1 schematisch ein Recherchensystem zur Ermittlung von Informatio-
 9 nen aus einer Datenbank,

10
 11 Fig. 2 eine konzeptionelle Darstellung des Recherchensystems nach Fig. 1,

12
 13 Fig. 3 eine schematische Aufstellung der im Recherchensystem nach
 14 Fig. 1 ablaufenden Prozesse,

15
 16 Fig. 4 eine alternative Darstellung für das Gesamtsystem nach Fig. 1,

17
 18 Fig. 5 schematisch die Struktur der Administration des Recherchensystems
 19 nach Fig. 1,

20
 21 Fig. 6 ein Schema für eine Anfrage an eine Mehrzahl von Suchmaschinen,

22
 23 Fig. 7 ein Schema für eine verteilte Abfrage und Zusammenführung der
 24 Ergebnisse,

25
 26 Fig. 8 ein Schema für verschiedene Suchmodi im Recherchensystem nach
 27 Fig. 1,

28
 29 Fig. 9 ein Schema für eine Filterung von Suchanfragen,

30
 31 Fig. 10 eine schematische Darstellung der im Recherchensystem nach
 32 Fig. 1 verwendeten Module,

Fig. 11 schematisch eine Darstellung der Struktur eines Filter-/Interpretermoduls,

Fig. 12 schematisch eine Struktur eines Integrationsmoduls,

Fig. 13 schematisch die Struktur eines Server-Engine-Moduls,

Fig. 14 schematisch ein dynamisches Model für einen Suchprozess,

Fig. 15 ein Ablaufschema für eine Initialisierung einer Server-Engine,

Fig. 16 ein Ablaufschema für die Bearbeitung einer Suchanfrage,

Fig. 17 ein Ablaufschema für die Vorbereitung von Adaptern für eine Suchanfrage, und

Fig. 18 ein Schema für eine Admin-Anwendung des Recherchensystems nach Fig. 1.

Das Recherchensystem 1 gemäß Fig. 1 ist zur Ermittlung von Informationen aus einer komplexen Datenbank, insbesondere aus dem Internet, dem World-Wide-Web oder einem anderen Index, mit besonders hoher Recherchenqualität ausgelegt. Dazu ist das Recherchensystem 1 datenseitig mit einer Mehrzahl von Suchmaschinen 2 verbunden, wie die derzeit zur gezielten Durchsuchung des Internets weit verbreitet im Einsatz sind.

Um dabei den gleichzeitigen und gemeinsamen Einsatz einer Mehrzahl von Suchmaschinen 2 zu ermöglichen, umfasst das Recherchensystem 1 in einer zentralen Auswerteeinheit 4 ein Integrationsmodul 6, dem die von den Suchmaschinen 2 gelieferten Trefferlisten oder Ergebnisdaten in Form von Ergebnisdatensätzen E zuführbar sind. Das Integrationsmodul 6 stellt dabei sicher, dass eine gemeinsa-

1 me, konsistent priorisierte Auswertung der von den verschiedenen Suchmaschi-
2 nen 2 gelieferten Ergebnisdatensätze E ermöglicht ist.

3
4 Dazu ist das Integrationsmodul 6 dazu ausgelagt, die eintreffende Ergebnisdaten-
5 sätze E zunächst in ein gemeinsames Datenformat zu konvertieren, so dass eine
6 konsolidierte Auswertung ermöglicht ist. Um dem Benutzer dabei eine gezielte
7 Auswertung der möglicherweise großen Anzahl von Informationen zu ermögli-
8 chen, ist die konsolidierte Priorisierung und Relevanzkennung der einzelnen Er-
9 gebnisdatensätze E vorgesehen, wobei nach vorgegebenen Relevanzkriterien
10 sämtlichen Ergebnisdatensätzen E untereinander vergleichbare Relevanzkenn-
11 werte zugeordnet werden. Um dies zu ermöglichen, ist das Integrationsmodul 6
mit einem Speichermodul 8 verbunden, in dem für die Zuweisung der Relevanz-
13 kennwerte relevante Priorisierungsparameter P hinterlegt sind.

14
15 Bei der Priorisierung kann insbesondere vorgesehen sein, die nach Vorgabe
16 durch einen Benutzer oder durch einen Betreiber ausgewählte Suchmaschinen 2
17 grundsätzlich mit einer vergleichsweise hohen Priorisierung und somit einem ver-
18 gleichsweise hohen Relevanzkennwert zu versehen. In diesem Fall erfolgt die
19 Wahl der Priorisierungsparameter abhängig von der jeweiligen Suchmaschine 2.
20 Alternativ oder zusätzlich kann vorgesehen sein, in der Art einer Normalisierung
21 die von einer Suchmaschine 2 gelieferten Ergebnisdatensätze E hinsichtlich ihres
22 Relevanzkennwerts zunächst auf denjenigen Ergebnisdatensatz E zu normieren,
24 der bereits von der jeweiligen Suchmaschine 2 aus mit dem höchsten Relevanz-
kennwert versehen wurde. In der Art einer prozentualen Darstellung könnte dies
25 somit bedeuten, dass der von der jeweiligen Suchmaschine 2 als am relevante-
26 sten erkannte Ergebnisdatensatz E mit einem vorläufigen Relevanzkennwert von
27 100 % versehen wird, wobei die anderen Ergebnisdatensätze E dieser Suchma-
28 schine 2 entsprechend ihrem von der Suchmaschine 2 zugewiesenen Aus-
29 gangsrelevanzkennwert einen entsprechend abgestaffelten Relevanzkennwert
30 erhalten. Wenn dies für sämtliche Suchmaschinen 2 gleichgewichtig durchgeführt
31 wird, erfolgt somit eine relative Priorisierung der Ergebnisdatensätze E unter-
32 schiedlicher Suchmaschinen 2 im Wesentlichen anhand ihrer relativen Relevanz

im Hinblick auf den von der jeweiligen Suchmaschine 2 als am relevantesten eingestuften Ergebnisdatensatz E.

Zusätzlich kann bei den Priorisierungsparametern P aber auch ein suchmaschinenspezifischer Kennwert hinterlegt sein, der allgemein oder abhängig von der jeweiligen Suchanfrage der jeweiligen Suchmaschine 2 einen Verlässlichkeitskennwert zuordnet.

Die Auswerteeinheit 4 umfasst weiterhin ein Distributionsmodul 10, das über einen Pufferspeicher 12 mit dem Integrationsmodul 6 kommuniziert. Das Distributionsmodul 10 ist weiterhin mit einem zweiten Speichermodul 14 verbunden, in dem für die Suchmaschinen 2 charakteristische Datensätze, insbesondere charakteristisch für deren Datenformat oder Syntax, hinterlegt sind. Eingangsseitig ist das Distributionsmodul 10 über ein Filtermodul 16 mit einem Ein-/Ausgabemodul 18 verbunden, das seinerseits eingangsseitig mit dem Integrationsmodul 6 verbunden ist. Das Ein-/Ausgabemodul 18 kann dabei insbesondere als herkömmliche Bedieneinheit, umfassend eine Tastatur und ein Ausgabemittel wie einen Bildschirm, ausgestaltet sein.

Bei der Bearbeitung einer vom Benutzer über das Ein-/Ausgabemodul 18 vorgegebenen Suchanfrage erfolgt zunächst eine Filterung dieser Anfrage im Filterbaustein 16. Dabei kann beispielsweise die Elimination von Suchanfragen eines bestimmten Kriterium vorgesehen sein oder es können beispielsweise länder- oder regionenspezifische Vorgaben für die Bearbeitung der Suchanfrage berücksichtigt werden. Des Weiteren erfolgt im Filtermodul 16 eine Vorauswahl der für die jeweiligen Suchanfragen als geeignet erkannten Suchmaschinen 2, so dass das Filtermodul 16 in der Art einer Grobstrategie bereits die mit der Suchanfrage endgültig zu betrauenden Suchmaschinen 2 mit vorgibt.

Die solchermaßen gefilterte Suchanfrage 2 wird anschließend dem Distributionsmodul 10 zugeführt, das unter Rückgriff auf die im zweiten Speichermodul hinterlegten suchmaschinenspezifischen Informationen und insbesondere Syntaxkenn-

1 werte die jeweilige Suchanfrage in eine Vielzahl individualisierter Suchanfragen in
2 den für die ausgewählten Suchmaschinen 2 passenden Datenformate umsetzt.
3 Anschließend gibt das Distributionsmodul 10 die individualisierten Suchanfragen
4 an die ausgewählten Suchmaschinen 2 weiter, wobei in der Art einer Protokollie-
5 rung und zur Verwaltungserleichterung im Pufferspeicher 12 ein Profil der Such-
6 anfrage gemeinsam mit einer diese identifizierenden Kennung hinterlegt wird.

7
8 Die durch die Zuführung der individualisierten Suchanfrage gestarteten Suchma-
9 schinen 2 absolvieren anschließend ihr Suchprogramm im World-Wide-Web und
10 stellen ihre Antwortsignale, beispielsweise in Form von Trefferlisten, in denen die
11 aufgefundenen Ergebnisdaten als Ergebnisdatensätze E nach der durch die je-
12 weilige Suchmaschine 2 erkannten Relevanz geordnet aufgelistet sind. Anhand
13 der charakteristischen Identifikationsdaten für die jeweilige Suchanfrage werden
14 die Ergebnisdatensätze E sodann im Integrationsmodul 6 der jeweiligen Suchan-
15 frage wieder zugeordnet, so dass eine gezielte Auswertung auch bei zeitlich sich
16 überschneidenden Suchvorgängen ermöglicht ist.

17
18 Im Integrationsmodul 6 erfolgt zudem, gegebenenfalls ebenfalls unter Rückgriff
19 auf die im zweiten Speichermodul 14 hinterlegten Syntaxinformationen, eine Kon-
20 vertierung der Ergebnisdatensätze E in das gemeinsame Datenformat. Anschlie-
21 ßend erfolgt unter Rückgriff auf die im Speichermodul 8 hinterlegten Priorisie-
22 rungsparameter P die Erzeugung eines Relevanzkennwerts für jeden Ergebnis-
23 datensatz E, wobei die solchermaßen erzeugten Relevanzkennwerte dazu dienen
24 sollen, eine konsolidierte und konsistente Ergebnishierarchie zwischen den Er-
25 gebnisdatensätzen E herzustellen. Im Ergebnis wird somit dem Benutzer eine
26 homogenisierte Ergebnis- oder Trefferliste erzeugt, bei der die als relevantest er-
27 kannten Ergebnisdatensätze E zuerst gelistet sind. Die Zuordnung des jeweiligen
28 Relevanzkennwerts unter Rückgriff auf die Priorisierungsparameter P kann dabei
29 insbesondere im Hinblick auf die jeweilige Suchmaschine 2 erfolgen, die den be-
30 troffenen Ergebnisdatensatz E geliefert hat. Alternativ oder zusätzlich kann bei der
31 Zuordnung des Relevanzkennwerts auch ein von der jeweiligen Suchmaschine 2
32 bereitgestellter Ausgangsrelevanzkennwert mit berücksichtigt werden.

1
 2 „ Zusätzlich zu den Suchmaschinen 2 ist das Recherchensystem 1 datenseitig mit
 3 einer Werbemitteldatenbank 20 verbunden. In der Werbemitteldatenbank, auch
 4 als „Ad-Server“ bezeichnet, sind in Form von so genannten „Banners“ Werbeele-
 5 mente hinterlegt, die von der Werbemitteldatenbank 20 in Reaktion auf eine ein-
 6 gehende Suchanfrage an das Recherchensystem 1 zurückgeliefert werden. Das
 7 Recherchensystem 1 ist dabei derart ausgestaltet, dass die Werbemitteldaten-
 8 bank 20 ausgewählte und spezifizierte Werbeelemente als Antwortdatensatz auf
 9 eine eingehende Suchanfrage zurücksendet, wobei die Auswahl und Spezifikation
 10 inhalts- oder themenbezogen angepasst an die jeweilige Suchanfrage erfolgt. Die
 11 Auswahl der zurückgesandten Werbeelemente erfolgt dabei dahingehend, dass
 12 das jeweilige Werbeelement inhaltlich oder thematisch einen Bezug zur Suchan-
 13 frage aufweist, so dass mit einem vergleichsweise erhöhtem Interesse des Benut-
 14 zers an dem jeweiligen Werbeelement gerechnet werden kann. Zu diesem Zweck
 15 übermittelt das Recherchensystem 1 die jeweilige Suchanfrage gemeinsam mit
 16 einem Kennungssignal an die Werbemitteldatenbank 20, wobei anhand des mit-
 17 gelieferten Kennungssignals die Auswahl möglicherweise für den Antwortdaten-
 18 satz geeigneter Werbeelemente vorgenommen wird.

19
 20 Fig. 2 verdeutlicht die beteiligten Komponenten und Aktoren im Bezug auf die
 21 (Sub)Systemgrenzen.

22 Rollenbeschreibung

24 Kunde

25 „Kunde“ ist jeder Benutzer, der eine Suchanfrage starten kann.

27 TSC Filtermanager

28 Die Rolle „TSC Filtermanager“ beschreibt den Mitarbeiterkreis, in dessen Verant-
 29 wortungsbereich die Filterdefinitionen liegen. Dieser Mitarbeiterkreis wertet außer-
 30 dem die Statistikprotokollierung des Recherchensystems aus.

TSC Admin

Die Rolle des „Admin“ beschreibt den Mitarbeiterkreis, der für den Betrieb des Recherchensystems verantwortlich ist.

Er administriert die Konfigurationsdatei (tsc.cfg) des Recherchensystems und überwacht die Fehlerprotokolle (tsc_error.log).

Beschreibung der externen Systeme

Inhalte

Kurzbeschreibung

Hier werden Suchergebnisse bereitgestellt, die auf private und Business-Homepage zeigen.

Schnittstelle

Die Schnittstelle ist analog zur WWW-Suche mit Fast.

P4P-Server

Kurzbeschreibung

Die Suchergebnisse des Pay4Performance(P4P)-Servers werden hier bereitgestellt.

Schnittstelle

Die Schnittstelle des GoTo-Suchcenters wird per HTTP-Request (inklusive Parameter) angesprochen. Die Ergebnisse werden in XML-Format geliefert.

FAST Suchcenter

Kurzbeschreibung

Hier werden internet-weite Suchergebnisse bereitgestellt, die mittels der Suchmaschine „AllTheWeb“ gefunden wurden.

Schnittstelle

Die Schnittstelle des FAST-Suchcenters wird per HTTP-Request (inklusive Parameter) angesprochen. Die Ergebnisse werden in XML-Format geliefert.

Ad-Server

Kurzbeschreibung

Der Ad-Server liefert entsprechend der Suchanfrage eine Werbe-Komponente, die in die Gesamtdarstellung der Suchergebnisse einfließt.

Recherchensystem Prozesse

In Fig. 3 werden die Prozesse des Recherchensystem in der Übersicht dargestellt und der Prozessdurchlauf im „Gut“-Fall beschrieben.

TSC Konfiguration

Die Konfiguration des Recherchensystem wird über Parameter gesteuert, die in einer eigenen Datei abgelegt sind. Diese Konfigurationsdatei wird im XML-Format im Dateisystem des Suchcenters abgelegt. Sie kann mit Hilfe eines Standard-Editors bearbeitet werden.

Die geänderte Konfiguration kann nur mit Hilfe des Restarts der „TSC Suche“ Prozesse erreicht werden.

Der Prozess „TSC Konfiguration“ beschreibt die notwendigen Arbeitsschritte für die Administration der Konfigurationsdatei des Recherchensystem.

Rolle	Aktivitäten
TSC Admin	<ol style="list-style-type: none"> 1. Konfiguration (tsc.cfg) editieren 2. Konfiguration (tsc.cfg) auf alle Produktionssysteme verteilen 3. „TSC Restart“ Prozess durchführen

TSC Ausschlussfilter pflegen

Die TSC Ausschlussfilter definieren Worte und Phrasen für die ein Suchstring, wenn sie in diesem enthalten sind, für bestimmte Kanäle gesperrt wird. Diese Sperrungen sind landesspezifisch und können zeitabhängig definiert werden. Gewöhnlich werden die Filter mit Hilfe einer eigenen Anwendung (der TSC Admin Anwendung) gepflegt. Letztendlich ist die Datei der TSC Ausschlussfilter das Produkt eines Exports aus der „Filter DB“ und liegt im CSV-Format vor. Diese Datei kann ggf. mit einem Standardeditor bearbeitet werden.

Die Aktivierung der geänderten Filter kann nur mit Hilfe des Restart der „TSC Suche“ Prozesse erreicht.

Folgende Arbeitsschritte werden für den Prozess „TSC Ausschlussfilter pflegen“ definiert:

Rolle	Aktivitäten
TSC Filtermanager	<ol style="list-style-type: none"> 1. Anmelden an der Anwendung „TSC Admin“ 2. Möglichkeit zur Ausführung folgender Aktionen: <ol style="list-style-type: none"> a. Filter Aktivieren b. Filter Erstellen c. Filter Ändern d. Filter Löschen 3. Aktivierung der geänderten Filter <ol style="list-style-type: none"> a. Filterdatei (tsc_filter_table) aus „Filter DB“ generieren b. Filterdatei (tsc_filter_table) auf alle Produktionssysteme verteilen c. „Sukzessive“ Neustart der „TSC Suche“ Prozesse

TSC Restart

Der „TSC Restart“ Prozess kann manuell oder automatisch durch andere Skripte, wie z. B. im Kontext der Filteraktivierung, gestartet werden.

Rolle	Aktivitäten
TSC Admin, TSC Filtermanager	<ol style="list-style-type: none"> 1. tsc_restart.sh starten <ol style="list-style-type: none"> a. Starten des tsc_restart.sh auf allen Produktionssystemen <ol style="list-style-type: none"> i. Ermitteln der PIDs aller Prozesse „TSC Suche“ ii. Sukzessive, zeitversetztes Senden des kill Signals an alle Prozesse

Die „fast-cgi“ Umgebung des Apache Web-Servers sorgt automatisch für das Neustarten der vorkonfigurierten Anzahl „TSC Suche“ Prozesse.

TSC Suchanfrage bearbeiten

Der Prozess „TSC Suchanfrage bearbeiten“ ist der zentrale Prozess, der durch das Recherchensystem realisiert wird.

Der folgende Ablauf skizziert diesen Prozesses auf hoher Ebene. Detailliert wird dieser Ablauf später im Dokument.

Rolle	Aktivitäten
Kunde	<ol style="list-style-type: none"> 1. Suche eingeben und abschicken
TSC Server Engine	<ol style="list-style-type: none"> 2. Suchanfrage entgegennehmen 3. Suchstring lesen 4. Attribute der Suchanfrage lesen
BRISBANE- Interpreter	<ol style="list-style-type: none"> 5. Suchstring in elementare Bestandteile zerlegen 6. Stopp-Wort-Filterung durchführen 7. Suchstring für die verschiedenen Kanäle aufbereiten

Rolle	Aktivitäten
BRISBANE- Integrator	<p>8. Suchanfrage je „Server“ starten</p> <p>9. Warten auf alle Ergebnisse der Suchanfrage (ggf. Timeout)</p> <p>10. Transformation der Ergebnisse (→ XML)</p> <p>11. Protokollierung der Statistikinformationen</p>
TSC Server Engine	12. Ergebnis zurückliefern .

Auswertung TSC Statistik

Die Statistikdatei kann mittels eines Texteditors eingesehen werden oder in andere Programme importiert werden, da die Datei im CSV-Format vorliegt.

Auswertung TSC Fehlerprotokoll

Mittels eines Texteditors kann das Fehlerprotokoll eingesehen werden.

Rolle	Aktivitäten
BRISBA NE-SC Admin	tbd.

TSC PlattformAnalysemodell

Use-Case-Diagramm, Grobanalyse

Fig. 4 beschreibt die groben Zusammenhänge der am Recherchensystem (TSC) beteiligten Akteure und Anwendungsfälle.

Use-Case: TSC Konfiguration

Der „TSC Admin“ erhält die Möglichkeit, die in einer XML-Datei abgelegte Konfiguration des Suchcenters einzusehen und zu verändern. Dazu bedient er das Admin-Tool, das auf einem separaten, nicht zur Produktivumgebung gehörenden Rechner läuft. Damit eine geänderte Konfiguration wirksam wird, ist ein Neustart der einzelnen Prozesse notwendig.

Business Process Diagramm

Die in Fig. 5 schematisch dargestellte Administration des Recherchensystems beinhaltet verschiedene Aktionen. Dazu zählen die Änderungen der Ausschlussfilter und das Aktivieren der Änderungen. Zur Aktivierung wird ein Script angestoßen, das die aktuelle Konfiguration auf die Rechner des Produktivsystems kopiert. Im Diagramm sind die Rechner des Produktivsystems an der unteren rechten Ecke zu sehen. Der Administrator selbst tritt als „worker“ auf und erscheint hier links oben im Diagramm.

Use-Case: TSC Ausschlussfilter pflegen

Der „TSC Filtermanager“ pflegt die Wörter, die von der Suche ausgeschlossen werden sollen. Diese Wörter können pro Land (Portal) als auch zeitlich eingeschränkt definiert werden.

Nach Änderung der Ausschlussfilter ist jeweils das Recherchensystem (TSC) neuzustarten, damit die Änderungen aktiv werden können.

Use-Case: Konfiguration und Filter aktivieren

Die Konfiguration wird mit Hilfe des Admin-Tools vom Admin-Rechner auf alle am Suchcenter beteiligten Rechner übertragen und dann ein Neustart des Systems durchgeführt. Durch einen Neustart werden u. a. neue Prozesse gestartet, die während ihrer Initialisierung die aktuelle Konfiguration und die Filtertabellen laden.

Use-Case: TSC Restart

Das Suchcenter wird mit Hilfe des Admin-Tools neu gestartet. Dabei werden alle Prozesse auf allen am Suchcenter beteiligten Rechnern geordnet beendet und neugestartet. Ein Neustart ist notwendig, damit Änderungen der Ausschlussfilter oder der Konfiguration wirksam werden.

Use-Case: Suchanfrage bearbeiten

Der „Kunde“ gibt in der Suchmaske des Ein-/Ausgabesystems seinen Suchausdruck ein. Dieser kann aus mehreren mit UND-, ODER- und NICHT-Operatoren verknüpften Wörtern bestehen. Dieser Suchausdruck wird nach Absenden an die Server-Engine von dieser interpretiert, an die Anforderungen der verschiedenen (z. T. externen) Suchmaschinen angepasst und an diese selbst geschickt. Unter Berücksichtigung von konfigurierbaren Timeouts werden die verschiedenen Suchergebnisse dann zu einer Gesamtansicht integriert und zum Browser des Kunden zurückgeliefert.

Use-Case: Auswertung Statistik

Der „TSC Filtermanager“ kann über einen Texteditor die Suchbegriffe und Verarbeitungszeiten einsehen. Zusätzlich sollen dem Auswerter Möglichkeiten bereitgestellt werden, die eine Ad-hoc-Standardauswertung nach noch zu definierenden Kriterien ermöglicht.

Use-Case: Auswertung Fehlerprotokoll

Der „TSC Admin“ kann über einen Texteditor die in einer Textdatei abgelegten Fehlermeldungen einsehen.

Use-Case-Diagramme, Feinanalyse

In Fig. 6 ist der Ablauf einer Anfrage an die Suchmaschinen schematisch dargestellt. „Interpretierte Anfrage senden“ bearbeitet den eintreffenden Suchauftrag,

1 passt diesen an die Syntax der beteiligten externen Suchmaschinen an und sen-
2 det ihn an diese weiter.

3 „Ergebnisse integrieren“ empfängt die Suchergebnisse der Suchmaschinen und
4 präsentiert sie dem Nutzer im gewünschten Format.

5
6 Fig. 7 zeigt schematisch den Ablauf einer verteilten Abfrage mit anschließendem
7 Zusammenführen der Ergebnisse. „Ergebnisse sammeln“ akkumuliert die von den
8 Suchmaschinen eintreffenden Suchergebnisse. Ein Suchergebnis muss innerhalb
9 eines konfigurierbaren Zeitlimits (Timeout) eintreffen. Ist dem nicht so, wird dies
10 als Fehler interpretiert.

11 „Errorhandling“ übernimmt das Protokollieren von Fehlerzuständen in eine Fehler-
datei.

12
13 „Priorisieren“ wichtet die von den verschiedenen Suchmaschinen gelieferten Er-
14 gebnisse nach bestimmten Geschäftsregeln.

15 Fehlerbehandelte und priorisierte Suchergebnisse werden „in Ergebnisseite inte-
16 grieren“ in das gewünschte Ausgabeformat überführt.

17
18 Der Nutzer/Kunde hat, wie in Fig. 8 dargestellt, zwei Möglichkeiten der Suche.
19 Bei „einfach Suchen“ kann der Nutzer einen oder mehrere Suchbegriffe und even-
20 tuell Suchoperatoren eingeben, die der FAST-CGI Syntax entsprechen müssen.

21
22 Bei „erweiterter Suche“ kann der Nutzer zuerst den Typ des gesuchten Webin-
halts auswählen (Webseite, Bild, Musik etc.). Je nach Inhaltstyp stehen dann an-
24 gepasste Suchoptionen zur Verfügung.

25
26 Der in Fig. 9 schematisch dargestellte Filter oder Interpreter bekommt zusätzlich
27 zu den Konfigurationsdaten den aktuellen „query string“. Diese Zeichenkette ent-
28 hält Parameter und die Suchanfrage des Nutzers. Die Suchanfrage wird entspre-
29 chend der Filterregeln des jeweiligen Systems (Suchmaschine) modifiziert. Ein
30 weiterer Schritt baut aus den jeweiligen „query strings“ und einigen Parametern
31 eine komplette URL für die Suchmaschine. Diese URL ist im o. g. Diagramm als
32 „Ziel_URL“ identifizierbar.

1 Modulüberblick

2 Das Gesamtsystem unterteilt sich, wie in Fig. 10 gezeigt, im wesentlichen in drei
3 Module. Diese werden im Folgenden näher beschrieben. In Interpreter und Inte-
4 grator existieren Adapterklassen, die jeweils auf genau ein Zielsystem zugeschnit-
5 ten sind. Alle Adapterklassen besitzen einen statischen und einen dynamischen
6 Teil. Der statische Teil wird durch Konfigurationsdaten repräsentiert. Der dynami-
7 sche Teil hängt von der jeweiligen Suchanfrage ab.

9 Modul Interpreter

10 Das Modul BRISBANE-Interpreter enthält, wie in Fig. 11 gezeigt, Klassen, die der
11 Filterung und sonstiger Aufbereitung der Suchanfragen dienen. Allgemein kann
12 man sagen, dass der Interpreter die Basisdaten für den Integrator bereitstellt. Die
13 Funktionalität der Filterung ist stark vom Zielsystem (Suchmaschine) abhängig
14 und verbirgt sich in den entsprechenden Adapterklassen.

16 Modul Integrator

17 Die in Fig. 12 gezeigten Klassen im Modul Integrator sind verantwortlich für Wei-
18 terleitung der modifizierten Suchanfragen an die Zielsysteme. Weitere Aufgaben
19 des Integrators sind das Zusammenstellen der Suchergebnisse und deren Kon-
20 vertierung von XML in HTML, das Darstellen der Ergebnisseite, sowie die Bereit-
21 stellung der Statistikdaten.

22
23 Für jede Anfrage an ein Zielsystem wird eine Instanz der Klasse Request erzeugt.
24 Diese Objekte stellen die Kommunikationsschnittstelle zwischen Suchcenter und
25 Zielsystemen dar.
26 Die Kommunikation mit den Zielsystemen wird durch den RequestManager ver-
27 waltet. Er ist jedoch nicht für die Darstellung der Ergebnisse verantwortlich.

Modul Server-Engine

Dieses in Fig. 13 anhand seiner Klassen gezeigte Modul repräsentiert den Hauptprozess des Suchcenters. Er besitzt eine Instanz der Klasse „Configuration“, die den Zugriff auf alle Konfigurationsdaten ermöglicht. Innerhalb des Prozesses werden Konfigurationsdaten in unterschiedlichen Strukturen wie z. B. AdapterMap abgelegt.

Die Server-Engine ist auch für das Instanziiieren von Interpreter und Integrator sowie das geordnete Beenden des Prozesses (bei Eintreffen eines Signals) verantwortlich.

Die Klasse Logging ist für die Protokollierung von Fehlern und Statistikdaten verantwortlich.

Dynamisches Modell

In Fig. 14 ist das dynamische Verhalten der einzelnen Komponenten und der Informationsfluss zwischen ihnen dargestellt.

Nach dem Start der Suchcenter-Software wird die Konfiguration ausgelesen. Je nach Konfiguration werden dann entsprechende Ausschlussfilter und eine länderspezifische XSL-Datei geladen. Der Prozess wartet nun auf eine Suchanfrage.

Wenn diese eintrifft, wird der Suchausdruck interpretiert. In Ausschlussfiltern vorkommende Wörter werden aus der Suchanfrage entfernt. Zusätzlich wird die Anfrage an die Syntax der beteiligten Suchmaschinen angepasst. Die modifizierten

Suchausdrücke werden nun an die Suchmaschinen gesendet. Abhängig vom Format der zurückgelieferten Ergebnisse ist eventuell eine Transformation in das HTML- Format notwendig (wenn beispielsweise die Ergebnisse im XML- Format vorliegen). Die Ergebnisse werden an den Browser des Kunden gesendet.

Der Prozess geht wieder in Warteposition, bis der nächste Request eintrifft.

Die einzelnen Suchanfragen werden mit Suchausdruck, Zeitstempel und Dauer der Verarbeitung in der Statistikdatei „tsc_statistics“ abgelegt. Zudem werden Fehlerzustände während des gesamten Verlaufs in der Datei „tsc_error_log“ eingetragen.

Modul Server Engine

Initialisierungsphase

Während der Initialisierung werden, wie in Fig. 15 gezeigt, abhängig von aktiven Kanälen, Konfigurationsobjekte und Instanzen aller Adapter erzeugt. Die Kanäle entsprechen den vorher angesprochenen Channels und repräsentieren ein ganz bestimmtes Zielsystem.

Verarbeitung einer Suchanfrage

Bei Eintreffen eines Requests werden, wie in Fig. 16 dargestellt, Instanzen von Interpreter und Integrator angelegt. Alle notwendigen Informationen für die Instanzen werden von der Server-Engine bereitgestellt. Der Interpreter bereitet nun die Modifikationen der Suchanfrage vor, der Integrator sendet diese an die Zielsysteme und liefert die formatierten Suchergebnisse zurück. Die Instanzen von Interpreter und Integrator werden anschließend zerstört.

Modul Interpreter

Der Interpreter veranlasst, wie in Fig. 17 gezeigt, alle für den Request benötigten Adapter, jeweils ein AdapterResult zu erzeugen. Jeder einzelne Adapter modifiziert die originale Suchanfrage unter Zuhilfenahme der entsprechenden Filterregeln. Diese Suchanfrage und weitere Parameter sind für das Erzeugen eines AdapterResult-Objektes notwendig. Das erzeugte AdapterResult-Objekt wird dem Interpreter zurückgegeben, der es einer Liste hinzufügt.

Modul Integrator

Nach Erzeugung der Integrator-Instanz befindet diese sich in einem initialen Zustand. Die Verarbeitung der Requests erfolgt durch den Aufruf der Methode „do-Requests()“.

1 Main des fast- Programms

2
3 Das Fast-CGI Programm enthält eine Funktion „main()“. Dieses wird vom Webserver
4 aufgerufen. Voraussetzung für das ordnungsgemäße Arbeiten des Programms
5 ist die korrekte Initialisierung und das Eintreten in die Warteschleife. Die Konfiguration
6 wird in einer Datei gehalten. Die Angabe des kompletten Pfades dieser Datei
7 erfolgt über eine Umgebungsvariable.

8 Konfiguration des Recherchensystem

9 Das Suchcenter wird über eine einzige Datei konfiguriert (tsc.cfg). Diese Datei befindet
10 sich auf einem separaten Rechner, der nicht als Produktivrechner verwendet wird. Die Konfigurationsdatei dient als Vorlage für alle Konfigurationsdateien
12 pro Rechner in der Produktionsumgebung. Bei Änderungen der Konfiguration
13 (z. B.: Modifikation des Verzeichnisses für die Statistikdateien) wird die Konfigurationsdatei
14 auf dem außerhalb des Produktivsystems befindlichen Rechner modifiziert und per Script auf alle Rechner im Produktivsystem kopiert.
15

16 Format der Filtertabelle

17 Auf Anforderung im Admintool wird aus der Datenbank (mysql) eine aktuelle Filterdatei
18 erstellt, die über die Verteilmechanismen an die Suchcenter-Rechner übertragen wird.
19

20 Diese Filterdatei hat zeilenweise Einträge, innerhalb einer Zeile sind die Werte
21 durch Semikolon getrennt.

22 Eine Zeile ist folgendermaßen sematisch aufgebaut (Werte in [] sind optional.):

23 Ländercode; zu filterndes Wort oder Phrase; [Datum Beginn]; Uhrzeit Beginn; [Datum Ende];
24 Uhrzeit Ende; Kanal 1; Kanal 2; ...;Kanal n

25 z. B.

26 DE;Neu;2000-10-10;12:00:00;2001-12-31;23:35:00;WWW;;;;;

27 DE;verboten;2001-10-20;01:00:00;2001-12-

28 31;20:00:00;WWW;P4P;BRISBANE;PHP;SP;AD

29 UK;Bier;;00:00:00;;00:00:00;WWW;;BRISBANE;;SP;

CH;Alkohol;2001-09-30;15:00:00;;00:00:00;WWW;;BRISBANE;PHP;;

Aktivierung von Konfiguration und Filtertabelle

Die Aktivierung von Konfiguration und Filtertabelle erfolgt durch das Starten von Scripts.

Format der Protokolldatei

Der Name der Statistikdatei wird in der Konfigurationsdatei (tsc.cfg) angegeben, z. B. tsc_statistics_<PID>. <PID> wird dann im Betrieb durch die Prozess-ID ersetzt. Da gleichzeitig mehrere Prozesse laufen, gibt es auch dementsprechend viele Protokolldateien.

Alle Statistikdateien liegen in einem zentralen Verzeichnis.

Diese Statistik-Dateien werden im CSV-Format gespeichert. Zur Auswertung existiert ein separater Prozess, der aufgrund des großen Datenvolumens auf einem eigens dafür eingerichteten Server laufen sollte. Die Such-Prozesse schreiben entweder ihre Daten direkt auf den Auswertungsserver oder der Auswertungsprozess sammelt in regelmäßigen Abständen die Statistik-Daten aller Such-Prozesse ein.

Aufbau der Datei

Die Statistik-Datei besteht aus 4 Grundspalten und pro realem Kanal kommen weitere 4 Spalten hinzu.

Grundspalten:

1. Eingegebene Suchbegriffe
2. Uhrzeit der Suche
3. Suchmodus
4. Gesamtzeit in ms oder Timeout

- Spalten pro Kanal:

1. Name
2. verwendete Suchbegriffe
3. Zeitdauer in ms oder Timeout
4. Anzahl der Ergebnisse

Mengengerüst

Grundspalten ca. 50 Byte

Kanalspalten je ca. 50 Byte

Maximallast 500 Anfragen pro Sekunden, ca. 250 Byte pro Anfrage:

→ 125 KByte pro Sekunde → 450 MByte pro Stunde → 10,8 GByte pro Tag

Format der Fehlerprotokollierung

Der Name der Fehlerprotokolldatei wird in der Konfigurationsdatei (tsc.cfg) angegeben, z. B. tsc_err_<PID>. <PID> wird dann im Betrieb durch die Prozess-ID ersetzt. Da gleichzeitig mehrere Prozesse laufen, gibt es auch dementsprechend viele Fehlerprotokolldateien.

Einträge in das Fehlerprotokoll geschehen zeilenweise. Innerhalb eines Eintrages sind die Werte kommasepariert und hat folgende sematische Bedeutung

Uhrzeit, Original-URL, angepasste URL, Zeit für den reinen Request, Anzahl

Treffer, Beschreibung des aufgetretenen Fehlers

z. B.:

12.10.2001 09:41:35,http://192.168.70.42:80/fast-

cgi/ubrpsc?lang=any&q=test&start=0&num=10&context=www&type=all&Port
alLan-

guage=de&context=www,http://66.77.74.21:8075/max_xml_b00?hits=10&q
uery=test&offset=0,HTTP-Call:145 ms,Total Hits:15737604,ERROR in

1 BrisbaneIntegrator::getQueryResults() 'ghttp_process () failed,Error:
2 getQueryResults

3
4 Wenn ein allgemeiner Fehler auftritt, wird dieser in der Form, Uhrzeit, Beschrei-
5 bung des aufgetretenen Fehlers protokolliert, z. B.

6 12.10.2001 09:41:35,ERROR in BrisbaneIntegrator::initXSLT() 'mpXMLStyleptr:
7 xmlParseMemory error',Error: can't initialize xslt-process. Process Exit

8 **Beschreibung der Stylesheets**

9 Die XSL-Stylesheets dienen innerhalb des Suchcenters dazu, die gesammelten
10 Antworten in XML-Form zu einer HTML-Seite entsprechend der aktuellen Layout-
vorgaben umzusetzen. Die Daten für eine Webseite werden dem Stylesheet im
12 XML, bzw. durch Stylesheetparameter übergeben. Pro Sprache existiert im Such-
13 center ein Stylesheet. Innerhalb des Stylesheets werden die Suchergebnisse in
14 HTML aufbereitet sowie die Links für das Vor- und Zurückblättern erzeugt. Ent-
15 sprechend den Voreinstellungen wird dann noch die entsprechende Reiterstruktur
16 über den Suchergebnissen und die Werbebanner bzw. die Formular für die neue
17 Suche erstellt.

18 **Administration der Filtertabelle**

19 **Kontext**

20 Der Kontext der Admin Anwendung ist in Fig. 18 dargestellt.

Zusammenfassung

Ein Recherchensystem und ein Verfahren zur Ermittlung von Informationen aus einer Datenbank, insbesondere aus dem World-Wide-Web, sollen zur Erzielung von gleichbleibend hohen Recherchequalitäten in besonderem Maße geeignet sein. Dazu ist erfindungsgemäß die gemeinsame und/oder gleichzeitige Nutzung einer Mehrzahl von Suchmaschinen (2) vorgesehen, wobei in einem Integrationsmodul (6), das die von den Suchmaschinen (2) gelieferten, jeweils eine Mehrzahl von Ergebnisdatensätzen (E) umfassenden Ergebnisdaten in ein gemeinsames Datenformat konvertiert, unter Rückgriff auf in einem Speichermodul (14) hinterlegte Priorisierungsparameter (P) die Zuordnung eines Relevanzkennwerts zu jedem Ergebnisdatensatz (E) vorgesehen ist.

Fig. 1

Bezugszeichenliste

1	Recherchensystem
2	Suchmaschinen
4	zentrale Auswerteeinheit
6	Integrationsmodul
8	Speichermodule
10	Distributionsmodul
12	Pufferspeicher
14	Speichermodule
16	Filtermodul
18	Ein-/Ausgabemodul
20	Werbemitteldatenbank
E	Ergebnisdatensätze
P	Priorisierungsparameter

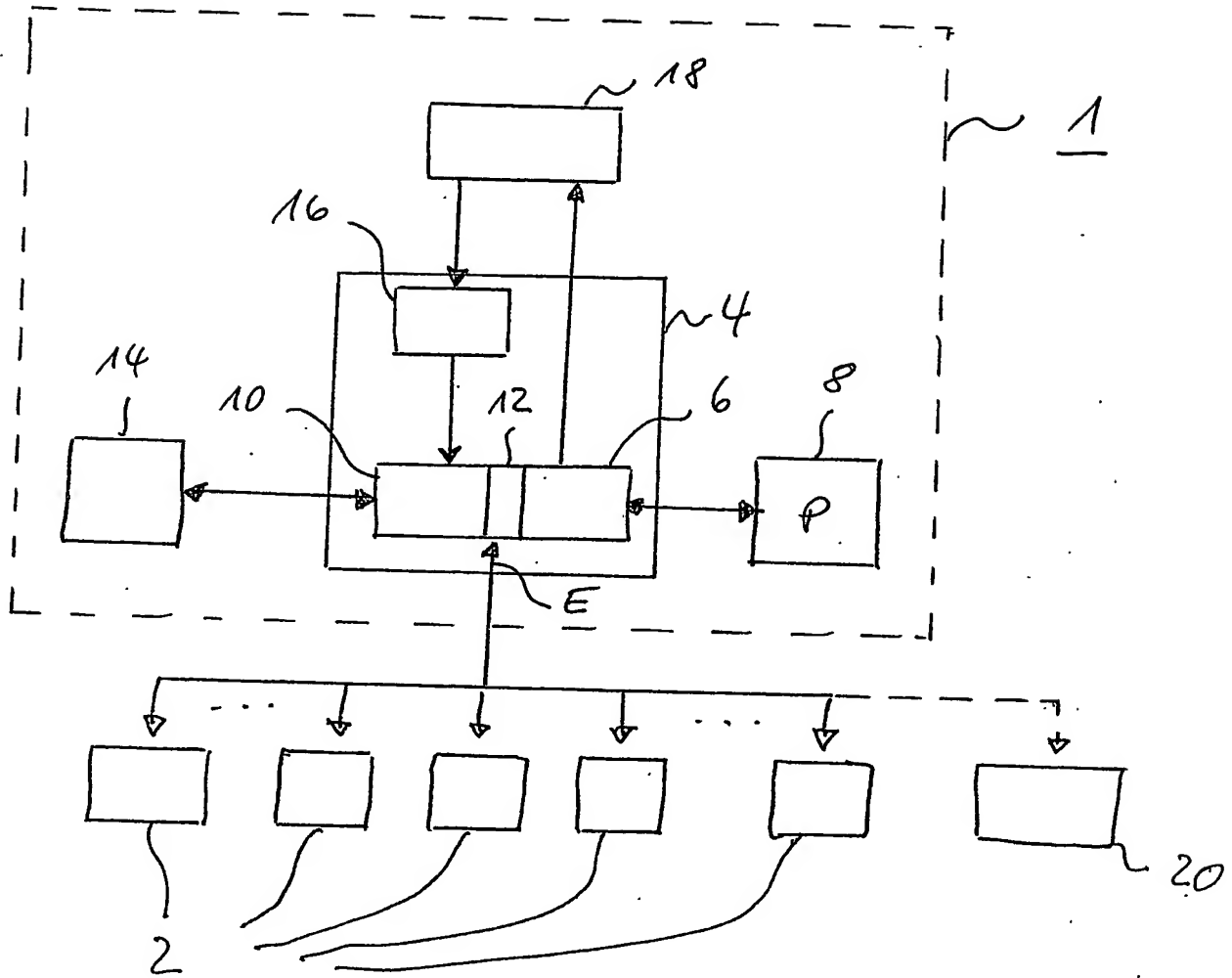


Fig. 1

2113

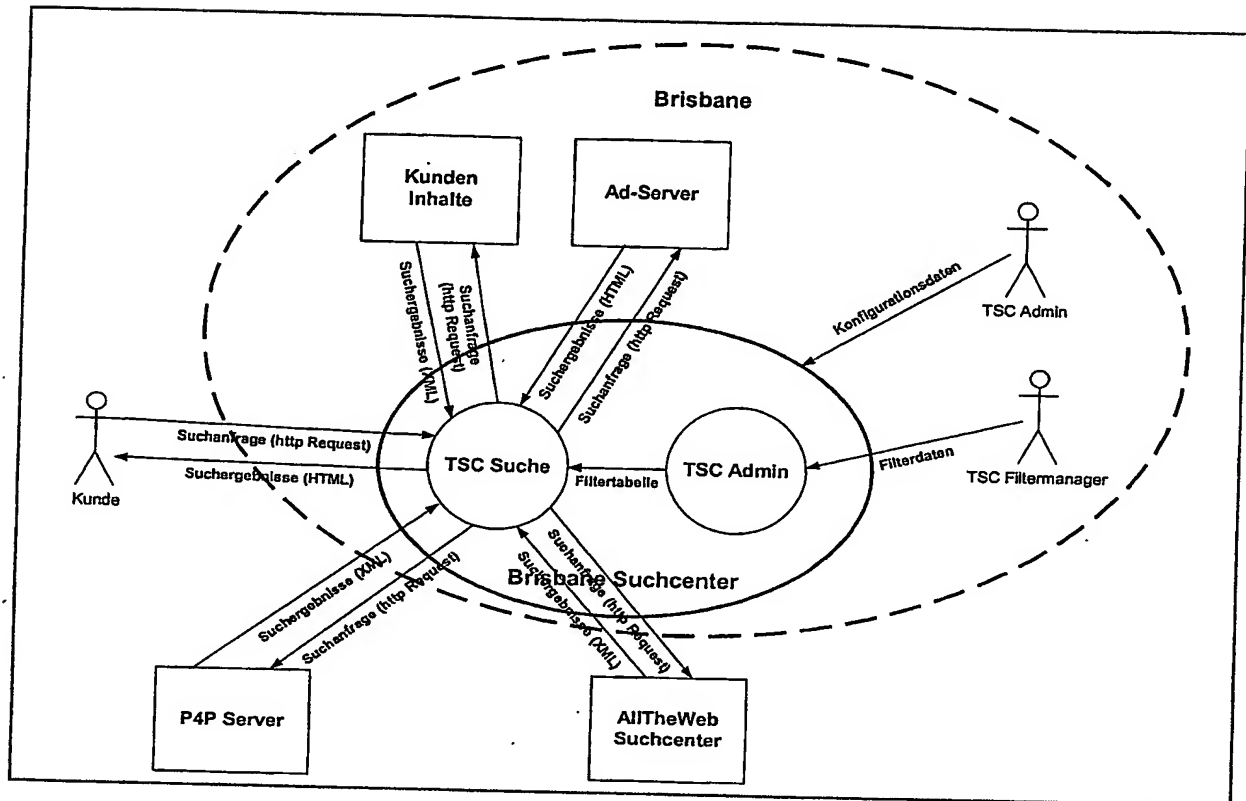


Fig. 2

3113

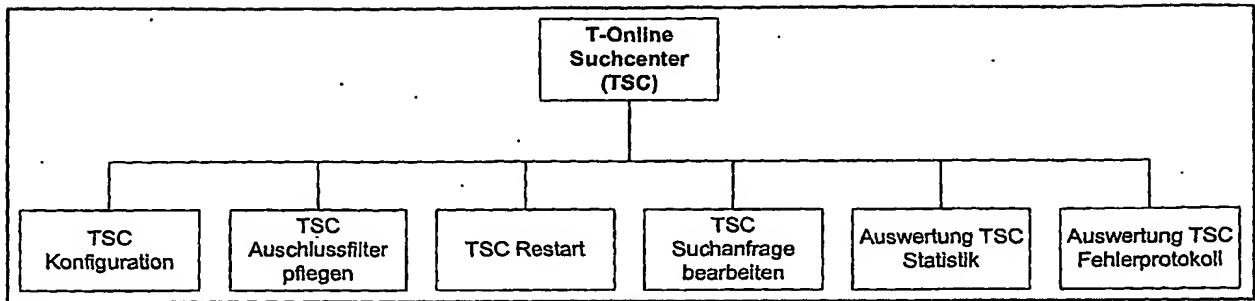


Fig. 3

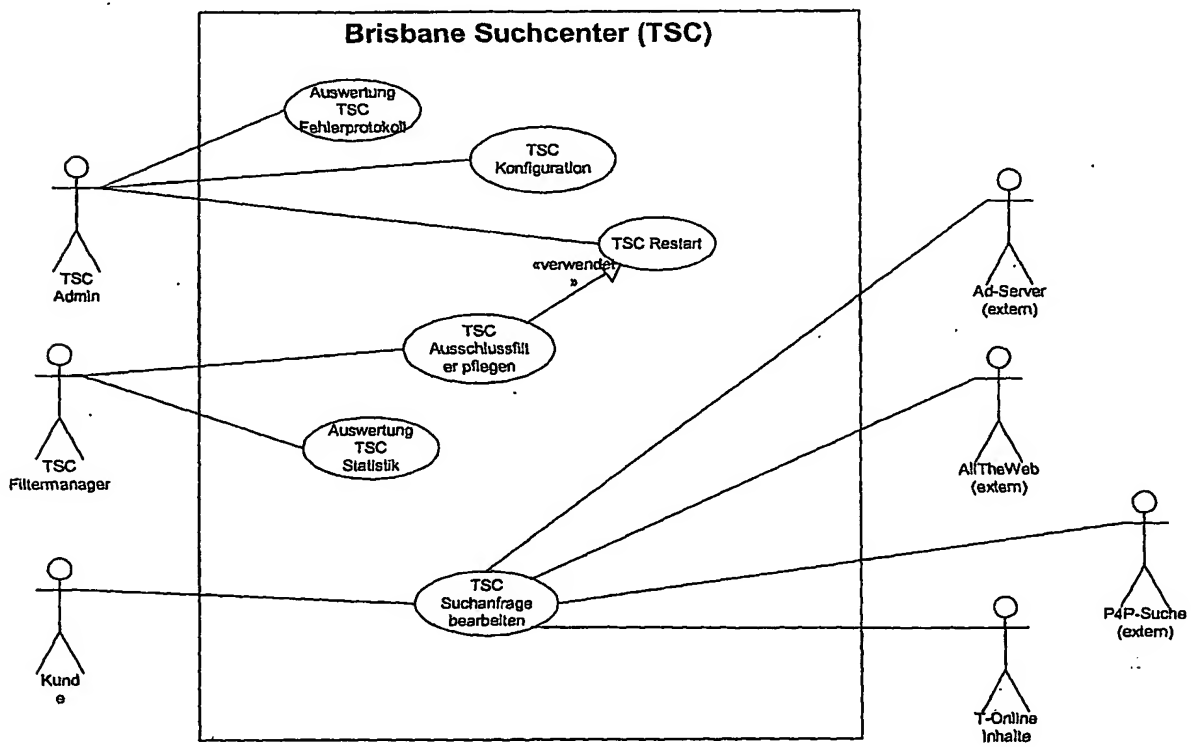


Fig. 4

4/13

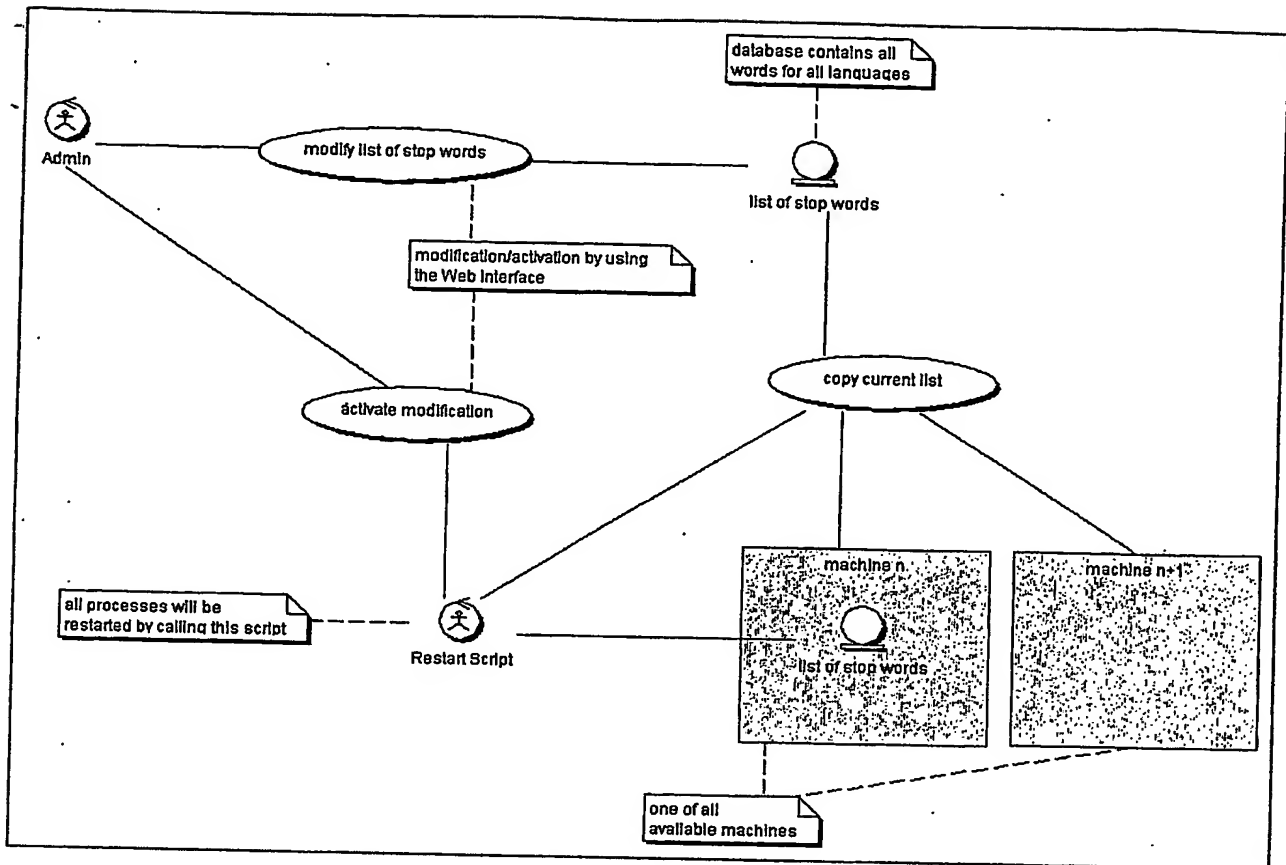


Fig.5

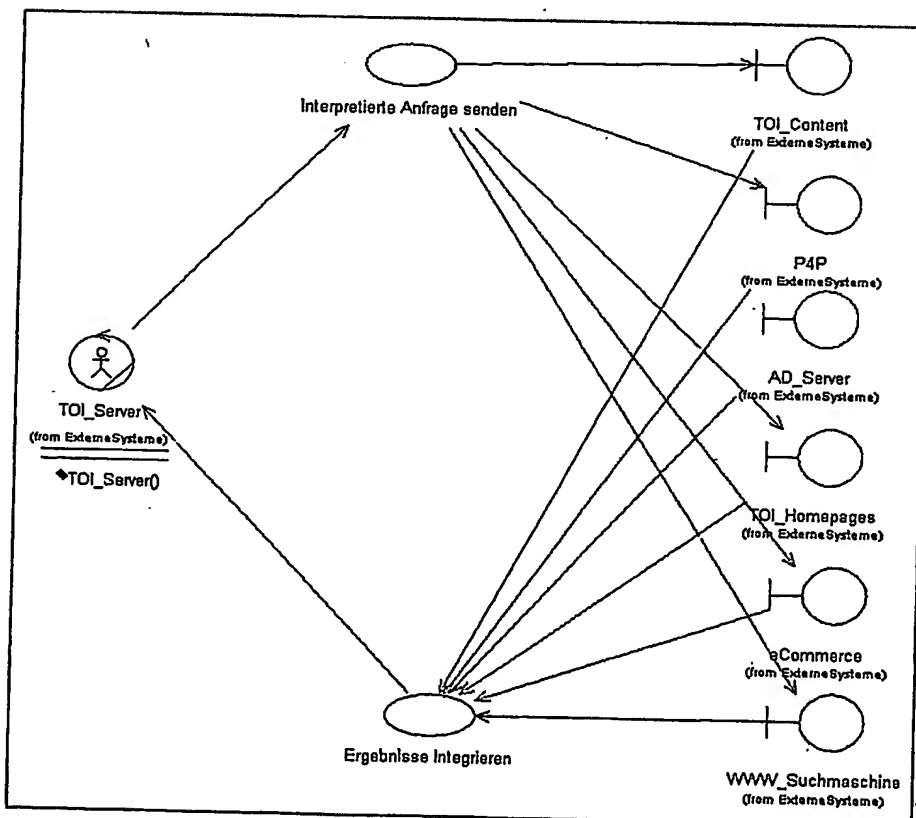


Fig.6

5/13

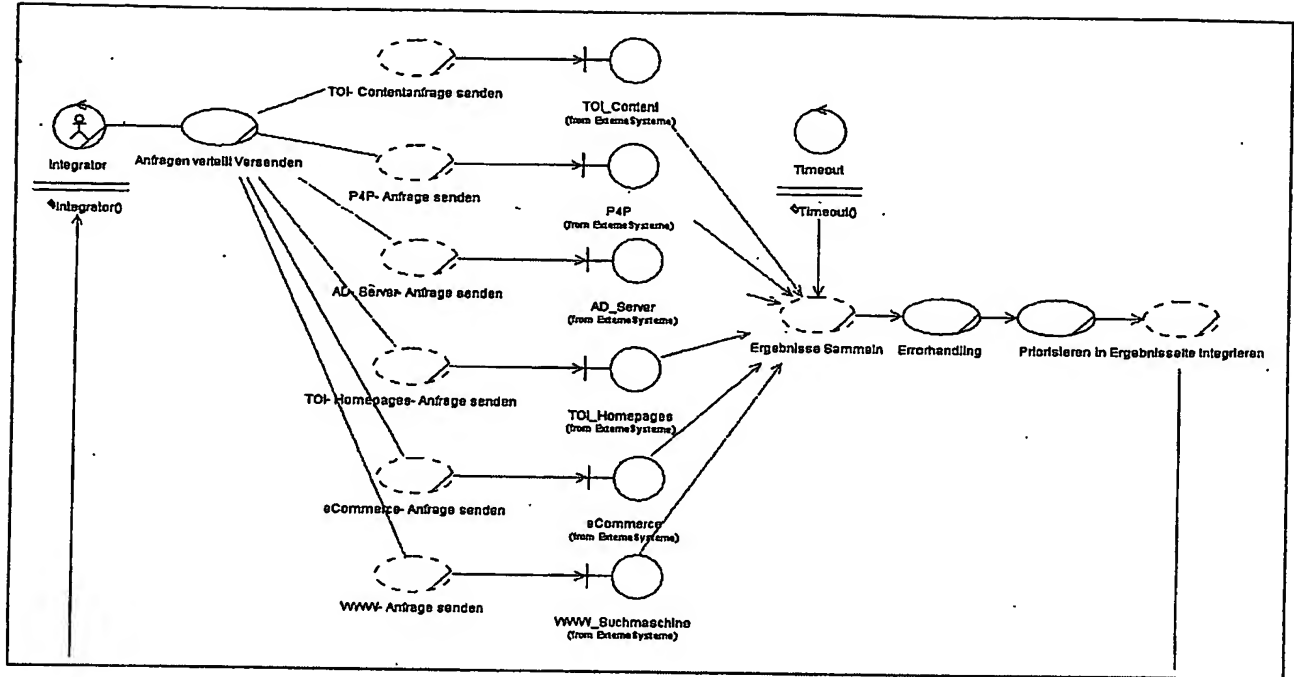


Fig. 7

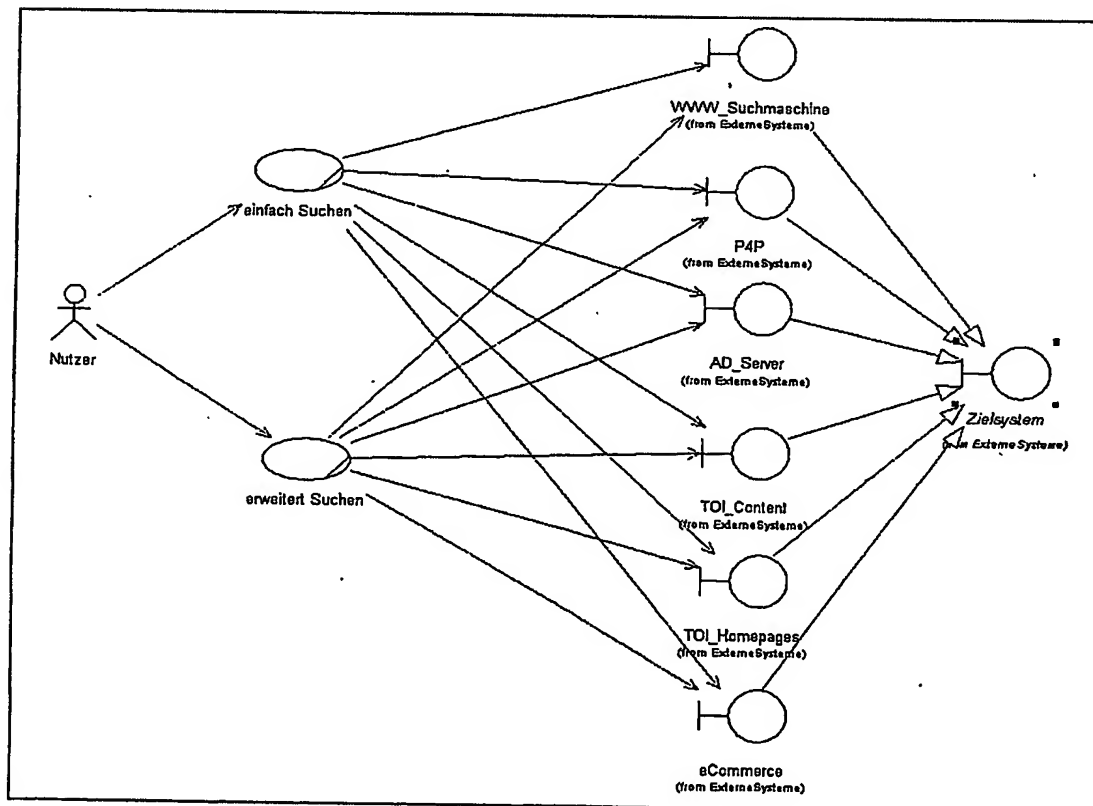


Fig. 8

6/13

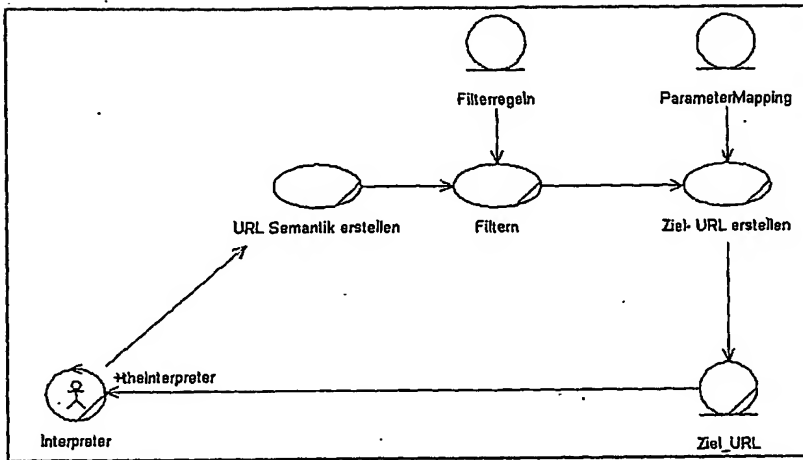


Fig. 9

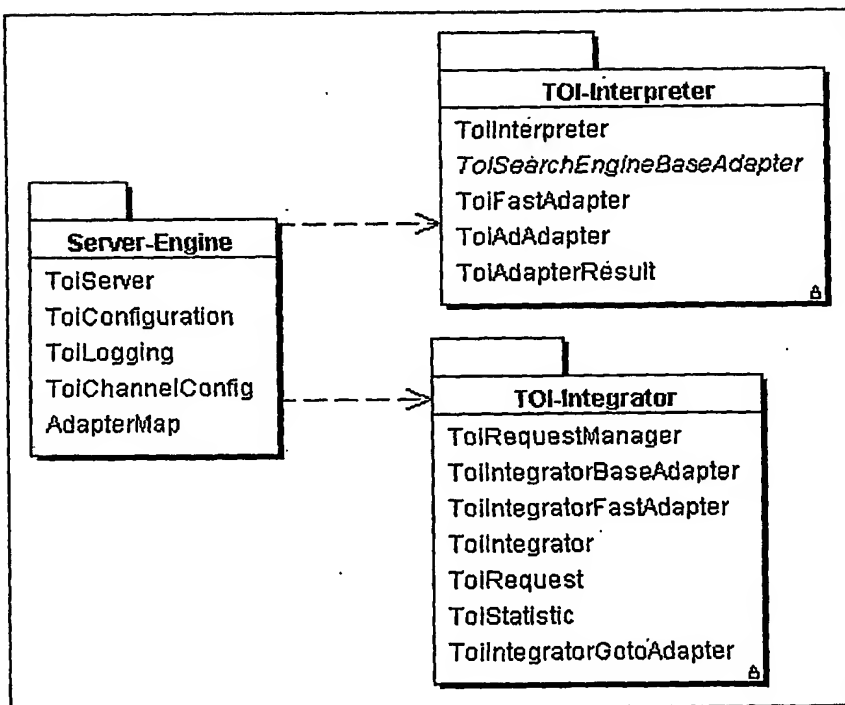


Fig. 10

4113

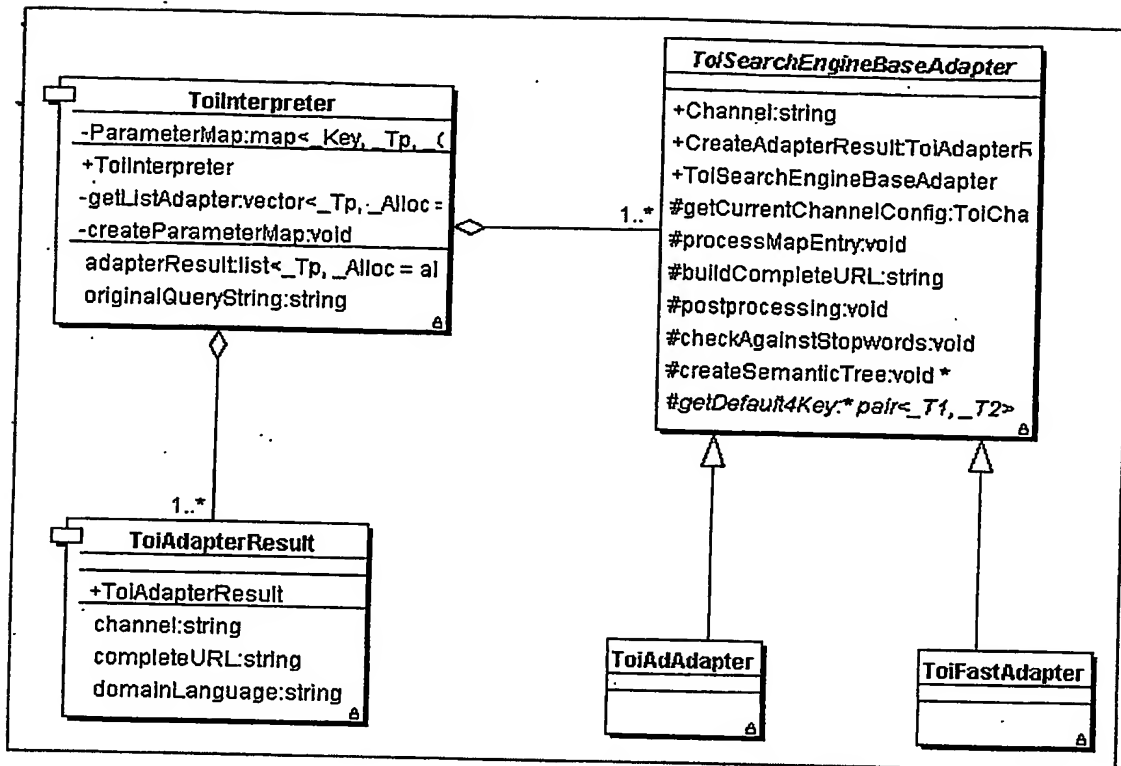


Fig. 11

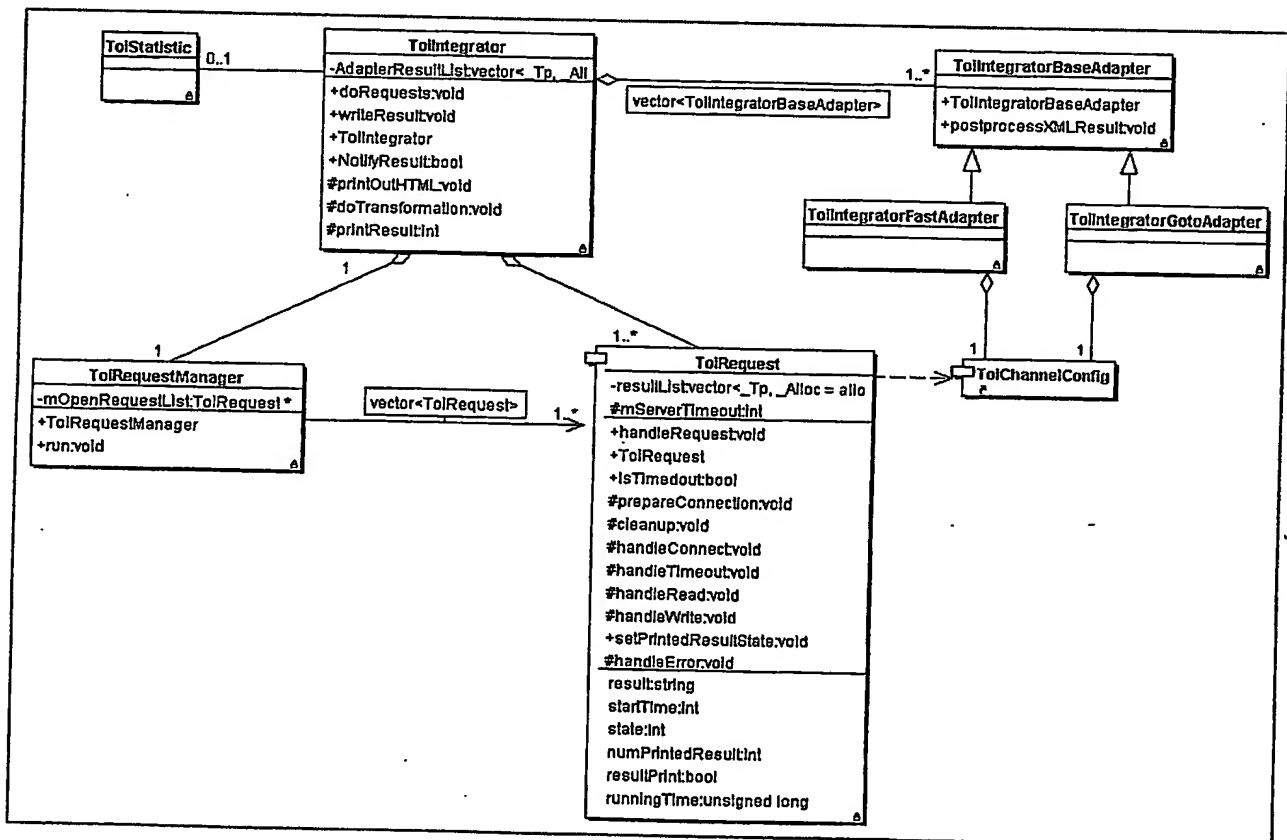


Fig. 12

8/13

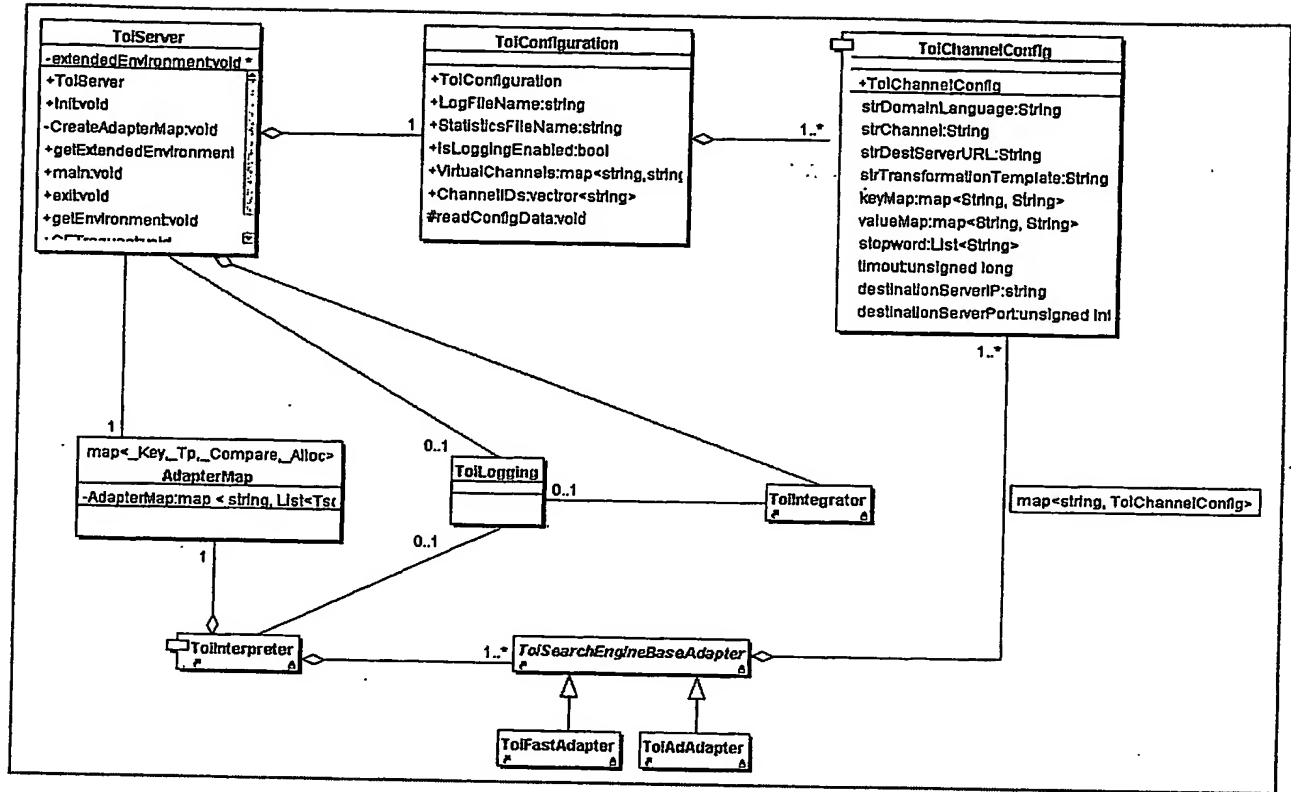


Fig. 13

9/13

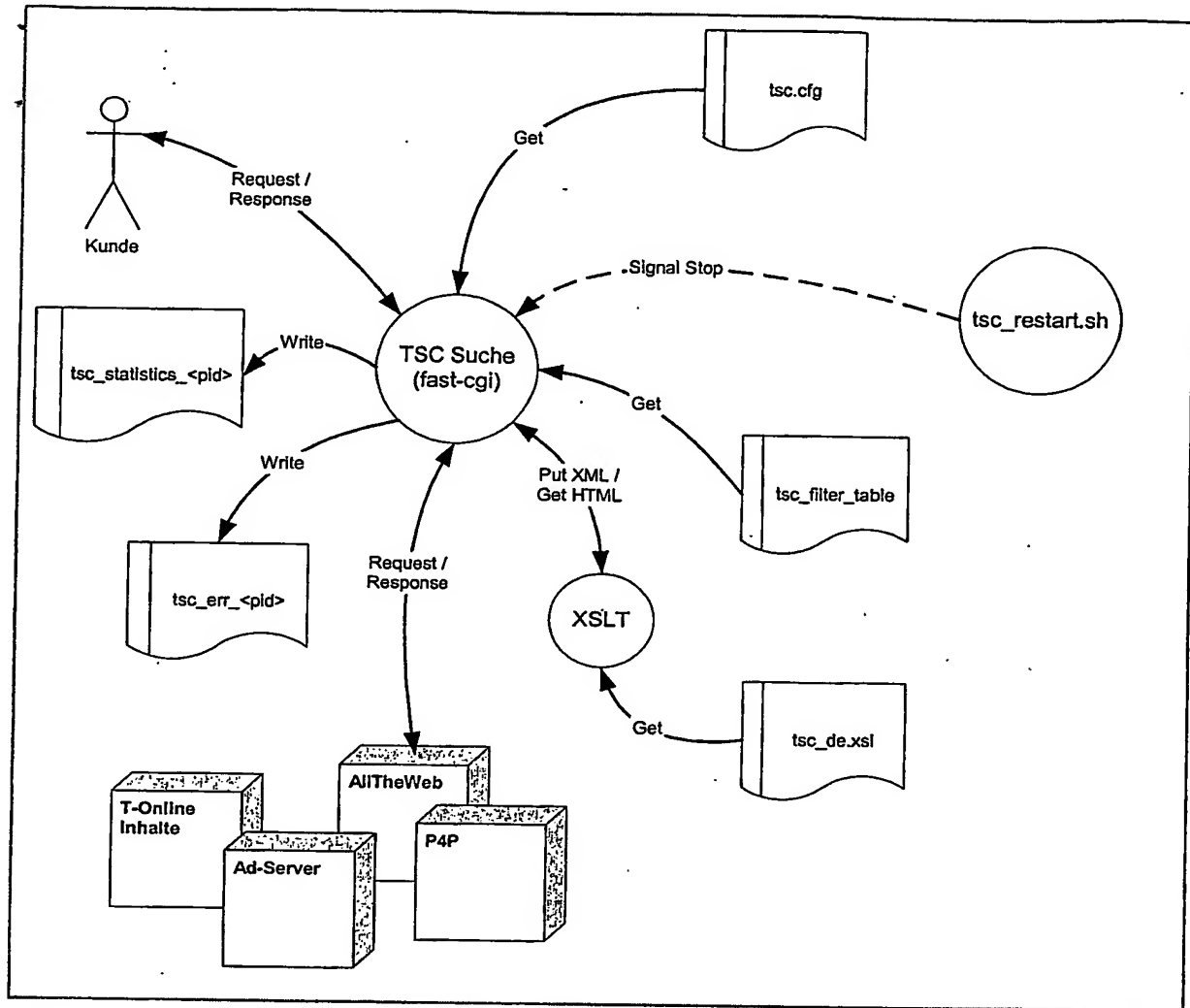


Fig. 14

10/13

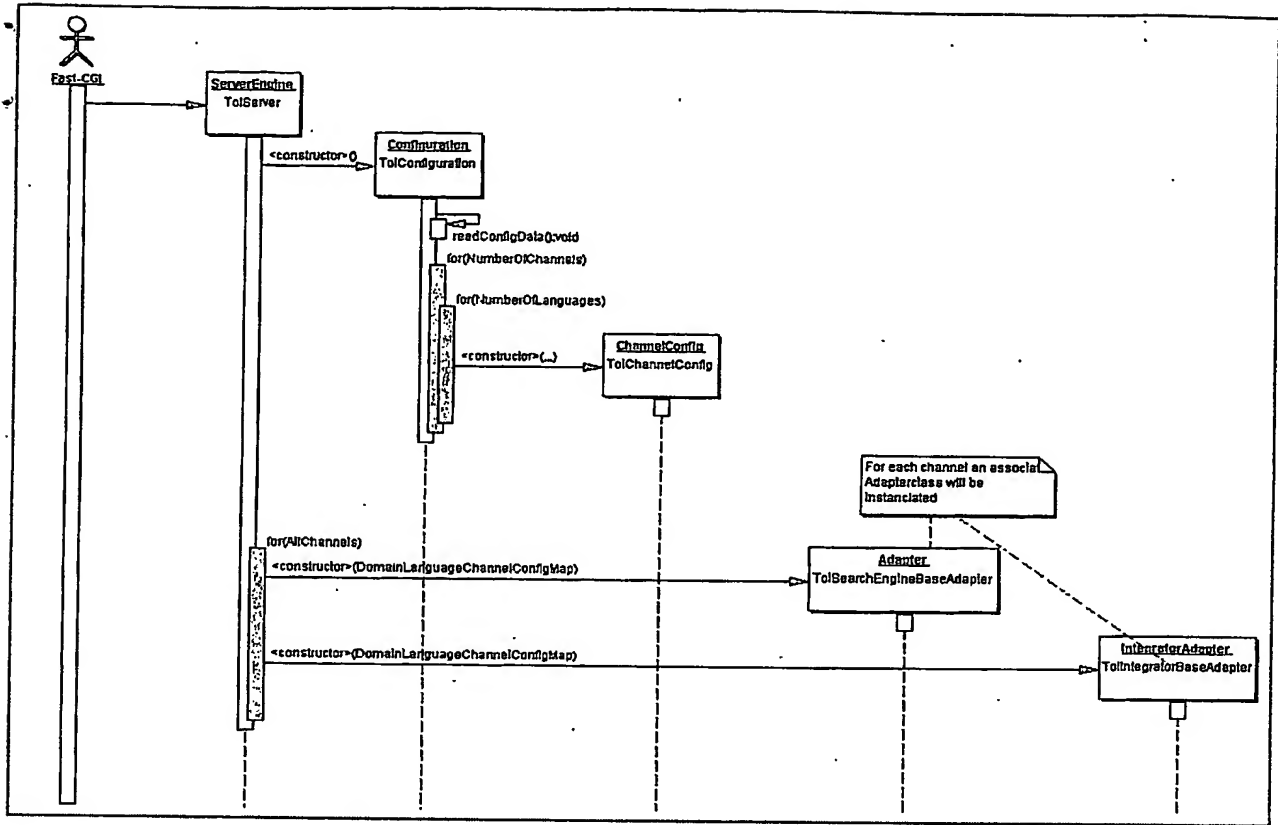


Fig. 15

11/13

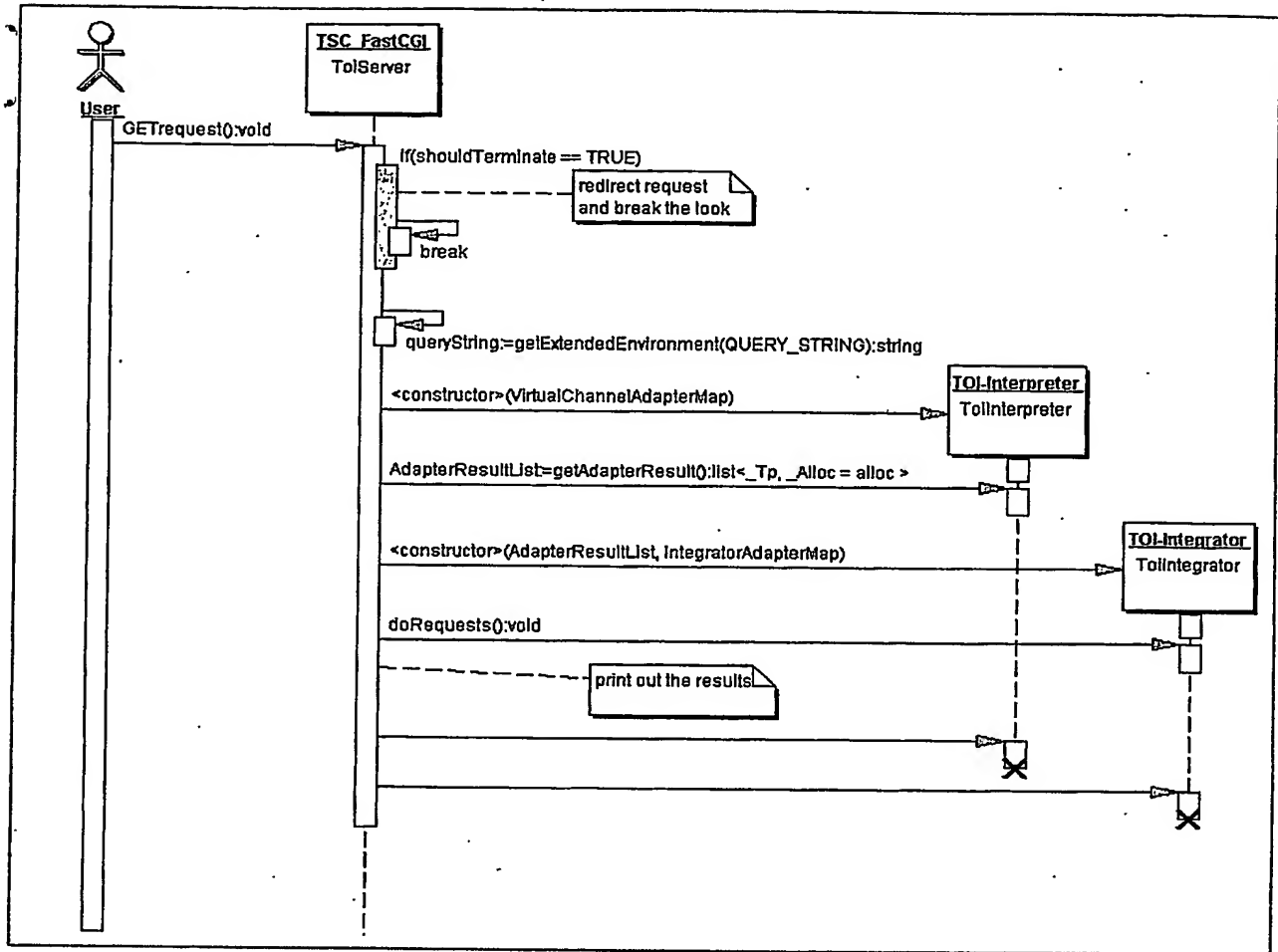
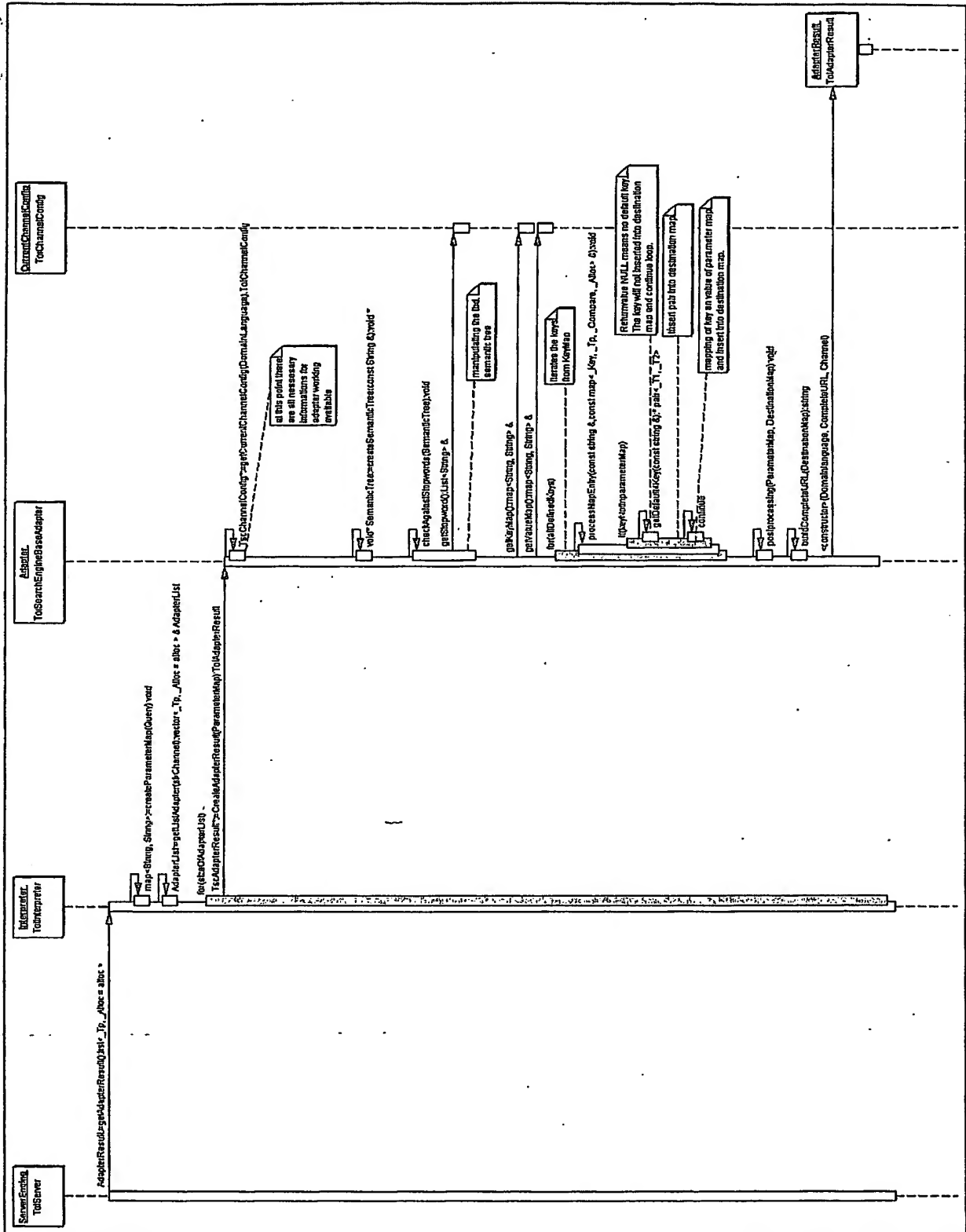


Fig. 16



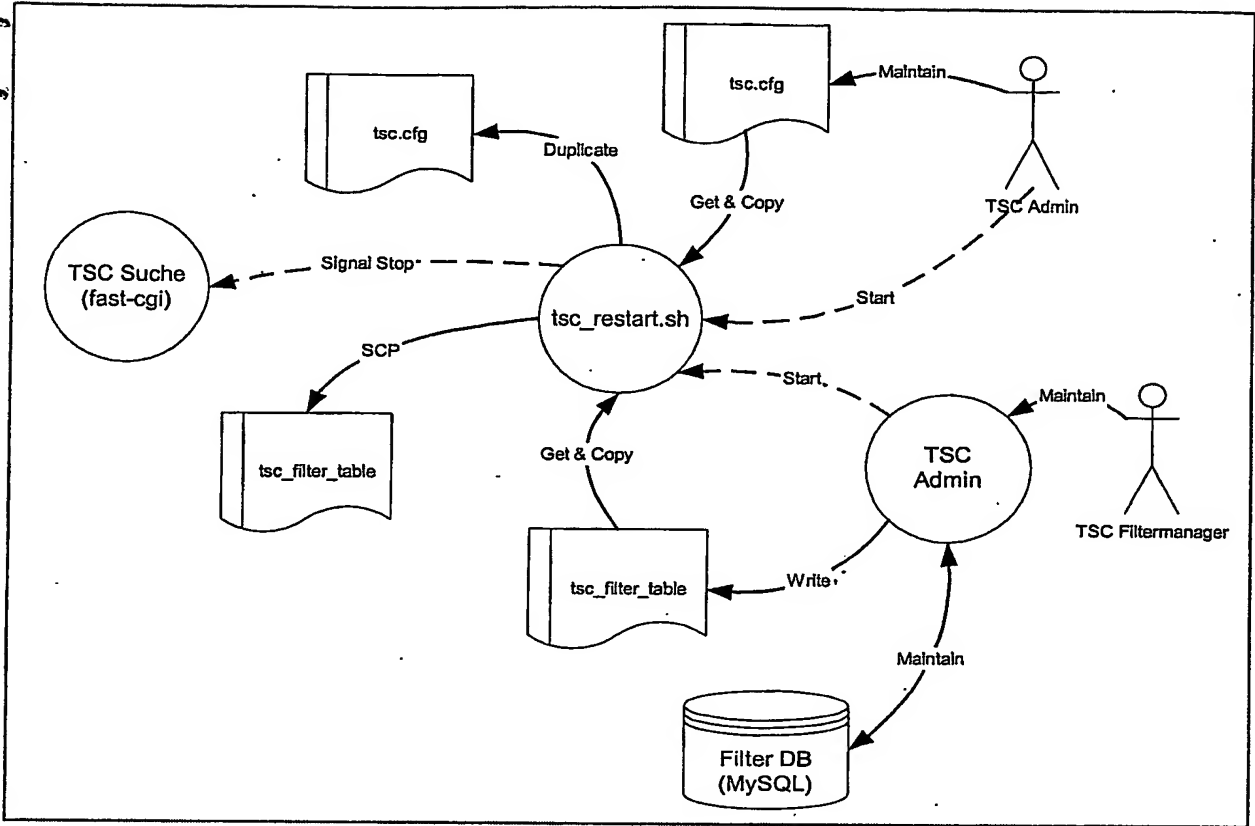


Fig. 18

This Page is inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURED OR ILLEGIBLE TEXT OR DRAWING
- ☒ SKEWED/SLANTED IMAGES
- ☐ COLORED OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REPERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images
problems checked, please do not report the
problems to the IFW Image Problem Mailbox**